AF 1635

#

#24/E.O.T.

1 mo.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
Before the Board of Patent Appeals and Interferences

In re Patent Application of

Atty Dkt. 117-347
C# M#

WILSON et al.

Group Art Unit: 1635

Serial No. 09/787,633

Examiner: Angell

Filed: July 10, 2001

Date: March 26, 2003

Title: TREATMENT OF INFECTION

OIPE JC34 MAR 26 2003 PATENT & TRADEMARK OFFICE

RECEIVED
MAR 31 2003
TECH CENTER 1600/2900

Assistant Commissioner for Patents
Washington, DC 20231

Sir:

☐ **Correspondence Address Indication Form Attached.**

☐ **NOTICE OF APPEAL**
Applicant hereby appeals to the Board of Appeals from the decision dated _____ of the Examiner twice/finally rejecting claims _____ ($ 0.00) $ 0.00

☒ An appeal **BRIEF** is attached in triplicate in the pending appeal of the above-identified application ($ 320.00) $ 320.00

☐ Credit for fees paid in prior appeal without decision on merits -$ ( )

☐ A reply brief is attached in triplicate under Rule 193(b) **(no fee)**

☒ Petition is hereby made to extend the current due date so as to cover the filing date of this paper and attachment(s) **($110.00/1 month**; $410.00/2 months; **$930.00/3 months**; $1450.00/4 months) $ 110.00

SUBTOTAL $ 430.00

☒ Applicant claims "Small entity" status, enter ½ of subtotal and subtract -$(215.00)
☐ "Small entity" statement attached.

SUBTOTAL $ 215.00

Less month extension previously paid on -$(0.00)

TOTAL FEE ENCLOSED $ 215.00

Any future submission requiring an extension of time is hereby stated to include a petition for such time extension. The Commissioner is hereby authorized to charge any deficiency, or credit any overpayment, in the fee(s) filed, or asserted to be filed, or which should have been filed herewith (or with any paper hereafter filed in this application by this firm) to our **Account No. 14-1140.** A duplicate copy of this sheet is attached.

1100 North Glebe Road, 8th Floor
Arlington, Virginia 22201-4714
Telephone: (703) 816-4000
Facsimile: (703) 816-4100
BJS:slj

NIXON & VANDERHYE P.C.
By Atty: B. J. Sadoff, Reg. No. 36,663

Signature: _____

03/27/2003 CCHAU1 00000001 09787633
02 FC:2251 55.00 OP

722930

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re Patent Application of

WILSON et al.

Atty. Ref.: 117-347

Serial No. 09/787,633

Group: 1635

Filed: July 10, 2001

Examiner: Angell

For: TREATMENT OF INFECTION

#25/Appeal Brief

* * * * * * * * * * * *

March 26, 2003

Assistant Commissioner for Patents
Washington, DC 20231

Sir:

APPEAL BRIEF

Pursuant to Rule 192, the appellants respectfully submit the present appeal, in triplicate.

(1) Real Party in Interest

The real party in interest is Medical Research Council, 20 Park Crescent, London W1N 4AL, United Kingdom, by way of an Assignment from the appellants which has been recorded in the U.S. Patent and Trademark Office on June 1, 2001, at Reel 011878, Frame 0113.

(2) Related appeals and interferences

There are no other appeals or interferences known to the appellants, the appellants' legal representative, or assignee which will directly affect or be directly affected by or have a bearing on the Board's decision in the pending appeal.

(3) Status of the claims

Application PCT/GB99/03180, which designated the U.S. was originally filed with 14 claims (published as WO 00/16758) which were replaced in the international phase with the 11 claims filed November 11, 2000, which were canceled in a Preliminary Amendment filed March 21, 2001, which added new claims 12-19. Claims 14-19 were canceled, and claim 12 was amended in an Amendment filed May 14, 2002. Independent claim 12 and dependent claim 13, which provide a screening method, are pending and are the subject of the present Appeal.

A copy of independent claim 12 and dependent claim 13, which are the subject of the present Appeal, is attached as Appendix A.

(4) Status of Amendments

There were no amendments filed in response to the final rejection of July 25, 2002 (Paper No. 17). A Response to Paper No. 17 was filed October 25, 2002. The Response filed October 25, 2002 has been considered by the Examiner and entered and overcome the Section 112, second paragraph, rejection of claims 12 and 13 stated on pages 2-3 of Paper No. 17. See, Advisory Action dated November 18, 2002 (Paper No. 22).

(5) Summary of Invention

The presently claimed invention provides a method of identifying or screening for a compound that inhibits the growth of an organism comprising the *ycf*24 gene. See, independent claim 12, lines 1-2, and, for example, page 2, line 28 through page 3, line 13 of the specification and original claims 4-6. The organism of dependent claim 13 is a malaria parasite . See, dependent claim 13, original claim 11, page 4, lines 24-26 and page 1, lines 12-14 of the specification.

The method of independent claim 12 entails at least the process steps of contacting a test compound with the *ycf*24 gene product and determining whether the test compound inhibits the activity of or binds to the product, wherein any such binding or inhibition suggests that the compound may inhibit the growth of the organism. See, independent claim 12, original claims 4-6, and page 2, line 28 through page 3, line 13 of the specification.

The presently claimed invention provides a method of screening or identifying compounds which may inhibit growth of an organism harboring the *ycf*24 gene, such as a malaria parasite. The appellants have discovered that the *ycf*24 gene is essential in organisms harboring the *ycf*24 gene and that loss or disruption of the *ycf*24 gene is lethal. See, page 2, lines 7-18 of the specification.

The claimed invention was exemplified, in part, with a malaria parasite (*Plasmodium falciparum* (ORF470)) *ycf*24 gene sequence (i.e., SEQ ID NO:1). See, page 6, lines 5-21 of the specification.

(6) Issue

The following one issue is the subject of the present appeal:

Whether the specification contains a written description of the invention of claims 12 and 13, as required by 35 U.S.C. §112, first paragraph.

(7) Grouping of the claims

The finally rejected claims do not stand or fall together. Pursuant to Rule 192(c)(7), the appellants will, in the argument section of this Brief, as required by Rule 192(c)(8), explain why the claims of the rejected group are believed to be separately patentable.

(8) Arguments

The specification contains a written description of the claimed invention. Consideration of the following and attached, and reversal of the Examiner's Section 112, first paragraph, "written description" rejection are requested.

The Examiner has appreciated that the claim recitation "the *ycf*24 gene" of line 2 of independent claim 12 is definite. See, page 1 of the Advisory Action dated November 18, 2002 (Paper No. 22), wherein the Examiner indicated the Section 112, second paragraph, rejection of claims 12 and 13 for reciting the same has been withdrawn in view of the appellants Response filed October 25, 2002.

In withdrawing the Section 112, second paragraph rejection of claims 12 and 13 for reciting "the *ycf 24* gene", the Examiner acknowledges that one of ordinary skill in the art will appreciate the metes and bounds of the claimed recitation.

The appellants urge the Board to appreciate that one of ordinary skill in the art will, for similar reasons, recognize that the appellants were in possession of the claimed invention, at least in so far as may be required by Section 112, first paragraph, as construed by the courts and Patent Office guidelines and further described herein.

The appellants have previously submitted the following five literature references with the Response of October 25, 2002[1] in which the authors describe *ycf 24* genes:

Kowallik *et al* (1995) Plant Molecular Biology Reporter, 13, 336-342;

Stirewalt *et al* (1995) Plant Molecular Biology Reporter, 13, 327-332;

Douglas and Penny (1999) J. Mol. Evol. 48, 236-244;

Reardon and Price (1995) Plant Molecular Biology Reporter, 13, 320-326; and

Denny *et al* (1998) Protist, 149, 51-59.

Further copies of these documents are attached as Appendix C.

The Board will appreciate that these documents provide a description of *ycf*24 gene sequences (i.e., compounds) from the following separate and distinct organisms:

[1] The appellants requested in their Response that the Examiner consider this evidence and acknowledge the same by returning the PTO 1449 Form listing the references which was attached to the Response or by returning a PTO 892 Form listing the same. The Examiner has not specifically indicated on the record that these references have been considered by either returning an initialed PTO 1449 Form or a PTO 892 Form listing the references. The Examiner has withdrawn the Section 112, second paragraph, rejection of claims 12 and 13, presumably after consideration of this evidence. The Examiner has also indicated in Paper No. 22 that the "REPLY FILED 22 October 2002" has been considered. These references were included with the Response filed October 25, 2002. The appellants are not aware of a Rely filed October 22, 2002 in the above. A copy of the appellants date-stamped post card receipt from the October 25, 2002 filing is attached as Appendix B. For completeness of the record, and to assure that the present appeal is considered in the most expeditious manner, the Examiner is requested in his Answer to correct and complete the record by specifically indicating that the Response filed October 25, 2002, including the evidence attached thereto, has been considered.

Odontella sinensis, *Cyanophora paradoxa*, *Guillardia theta*, *P. purpurea*, and Apicomplexans.

Specifically, Kowallik *et al* describes the chlroplast geneome of a chlorophyll-containing alga, *Odontella sinensis* and reports that one of the genes in the genome is "*ycf 24* gene" (see Figure 1, page 337, second line from the bottom and page 340). Stirewalt *et al* describes the nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa* and reports that the genome contains *ycf 24* (see the fourth row in page 329). Douglas and Penny describes the complete sequence of the plastid genome of the cryptophyte alga, *Guillardia theta*, and reports that it contains *ycf 24*. Douglas and Penny also confirms that *ycf 24* has been identified in other photosynthetic lineages (see Table 1 on page 239). Reardon and Price is a review article about the sequencing of plastid genomes of non-green algae and confirms that *ycf 24* has been recognized in such genomes (see the 5th item on page 326, i.e., "*P. purpurea*, *C. paradoxa*, *O. sinensis*, others"). Denny *et al* discusses the evidence for a single origin of the 35kB plastid DNA in Apicomplexans and notes that the *ycf 24* gene is highly conserved in the plastid across the different species (see page 53, right column, under the heading "The ORF470 Region").

The above-cited and attached references demonstrate that persons skilled in the art consider and considered the term "*ycf*24 gene" to be clear enough to be used in published scientific articles. They also show that the "*ycf 24* gene" has a highly conserved sequence (i.e., compounds) across species and that this allowed persons skilled in the art to recognize the gene (i.e., compounds) in a variety of genomes; the

references demonstrate that, when a genome from a new species was sequenced, any *ycf*24 gene was readily identified by its sequence.

The appellants are not claiming in the present application any specific *ycf*24 gene sequence but rather the use of a *ycf*24 gene sequence or a *ycf*24 gene product in a method based on the appellants discovery of the activity of the *ycf*24 gene or *ycf*24 gene product. The appellants have exemplified the claimed method in the present disclosure by providing *ycf*24 sequences of *Plasmodium falciparum* (SEQ ID NO:1), *Synechocystis* PCC6803 (SEQ ID NO:2), and *Escherichia coli* (SEQ ID NO:3). These exemplified *ycf*24 gene sequences, taken with the recognized *ycf*24 gene sequences of the art, and the whole of the present disclosure, will lead one of ordinary skill in the art to appreciate that the appellants were in possession of the claimed invention at the time the application was filed.

As further evidence of known *ycf*24 gene sequences (i.e., compounds), the appellants attached, as an example of the advanced level of skill in the art, as Appendix D, a copy of twenty six (26) *ycf*24 gene sequences obtained from the public NCBI database (i.e.,

http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?CMD=search&DB=nucleotide

Specifically, *ycf*24 gene product sequences from the following organism and/or clones are provided in the attached Appendix D[2]:

> Synechocystis sp. PCC 6803 DNA,
> Thermosynechococcus elongatus BP-1,
> Nostoc sp. PCC 7120

[2] Where the indicated data base entry contained more than the *ycf*24 gene and/or gene product, such as the complete genomic sequence, the appellants have provided in the attached, for conservation of paper and space, the sequence obtained by a word search of the term "ycf24" within the retrieved database record. The appellants would be happy to provide a complete copy of the retrieved database record upon the further request of the Examiner and/or the Board.

Synechocystis sp. PCC 6803
Methanobacterium thermoautotrophicum str. Delta H,
Toxoplasma gondii apicoplast,
Odontella sinensis chloroplast,
Cyanophora paradoxa cyanelle,
Porphyra purpurea chloroplast,
Guillardia theta chloroplast,
Cyanidium caldarium chloroplast,
Thermosynechococcus elongatus BP-1 section 2/9
Methanobacterium thermoautotrophicum from bases 1050856 to 1062059 (section 90 of 148),
Nostoc sp. PCC 7120 DNA, section 9/19,
Odontella sinensis complete chloroplast genome,
Skeletonema costatum chromoplast ycf24 gene, partial,
Cyanidium caldarium strain RK1 chloroplast,
Toxoplasma gondii chloroplast,
Guillardia theta complete plastid genome,
Porphyra purpurea chloroplast,
E.coli genomic DNA, Kohara clone #321(38.1-38.4 min.),
E.coli genomic DNA, Kohara clone #320(37.9-38.3 min.) and
Cyanophora paradoxa cyanelle.

The following *ycf*24 gene products are specifically provided from this database (the date of the deposit is indicated in parentheses after the genomic deposit accession number):

Synechocystis sp. PCC 6803 DNA gi:1001701 (1995)
MSSTTVKNLVNQPYKYGFVTNIEADAIPRGLSEDVVRLISAKKNEPEFMLDFRLRAYRHWLTMAEPTWPA
VHYPPIDYQDIIYYSAPKQSKKKLESLDEVDPALLETFEKLGIPLSEQKRLSNVAVDAIFDSVSIGTTFKEKLAEDG
VIFCSISEALQEHPDLVQKYLGSVVPTADNFFAALNSAVFSDGSFVFIPKGVKCPMELSTYFRINNGDTGQFERTLI
IAEEGASVSYLEGCTAPMYDTNQLHAAVVELVALDNADIKYSTVQNWYAGDENGKGGIYNFVTKRGLCKGVNSKISW
TQVETGSAITWKYPSCVLVGDNSVGEFYSIALTNNKQQADTGTKMIHIGKNTRSIIISKGISAGNSANSYRGLVKMG
PKAQGARNYSQCDSMLIGDRAAANTFPYIQVDNNTAKVEHEASTSKIGEDQLFYFAQRGISEEDAVSMLVSGFCKDV
LNELPMEFAAEADKLLSLKLEGTVG

Thermosynechococcus elongatus BP-1 gi:22297544 (2002)
MSATVQSLVNQPYKYGFVTPIETETIPKGLNEDIIRLISAKKNEPEFMLEFRLRAYRQWLKMSEPQWPRV
SYPPINYQDIVYYSAPKQKEKLKSLDEVDPVLLETFEKLGIPLSEQKRLTNVAVDAIFDSVSVATTFREELAKQGII
FCSISEALQDYPELVQKYLGSVVPIGDNFYAALNSAVFSDGSFVYVPKNTRCPMELSTYFRINNGESGQFERTLIIA
DAGSYVSYLEGCTAPMFDTNQLHAAVVELVALDNAEIKYSTVQNWYAGDENGKGGIYNFVTKRGLCLGRNSKISWTQ
VETGSAITWKYPSCVLVGDNSVGEFYSVALTNHYQQADTGTKMIHIGKNTRSRIVSKGISAGHSQNSYRGLVKIGPK
ATGARNYSQCDSMLIGDTAAANTFPYIQVQNPTAQVEHEASTSKIGEDQLFYFAQRGISAEDAVSMMISGFCRDVFN
QLPMEFAVEADRLLSLKLEGSVG

Nostoc sp. PCC 7120 gi:17227497 (2001)
MSATVKTLVNQPYKYGFVTDIEADTIPRGLDEDVVRLISTKKNEPEFMLEFRLRAFRQWQKMTEPTWPSV
KYPPIDYQNIIYYSAPKQKKAKLNSLDEVDPTLIETFEKLGIPLSEQKRLANVAVDAIFDSVSVATTFKEKLAKDGV
IFCSISEALQEHPELIKKYLGSVVPIADNYFAALNAAVFSDGSFVYIPKGVKCPMELSTYFRINSGDTGQFERTLIV

AEEGSYVSYLEGCTAPMYDSNQLHAAVVELVALDNAEIKYSTVQNWYAGDANGKGGIYNFVTKRGLCQGVNSKISWT
QVETGSAITWKYPSCVLVGDNSVGEFYSVALTNNMQQADTGTKMIHIGKNTRSTIISKGISAGQSSNSYRGLVKINP
TAKGARNYSQCDSMLIGDNAHANTFPYIQVQNNTGKVEHEASTSKIGEDQLFFFAQRGISSEDAISMMISGFCKDVF
NQLPMEFAVEADKLLSLKLEGSVG

Synechocystis sp. PCC 6803 gi:16329170 (1995)

MSSTTVKNLVNQPYKYGFVTNIEADAIPRGLSEDVVRLISAKKNEPEFMLDFRLRAYRHWLTMAEPTWPA
VHYPPIDYQDIIYYSAPKQSKKKLESLDEVDPALLETFEKLGIPLSEQKRLSNVAVDAIFDSVSIGTTFKEKLAEDG
VIFCSISEALQEHPDLVQKYLGSVVPTADNFFAALNSAVFSDGSFVFIPKGVKCPMELSTYFRINNGDTGQFERTLI
IAEEGASVSYLEGCTAPMYDTNQLHAAVVELVALDNADIKYSTVQNWYAGDENGKGGIYNFVTKRGLCKGVNSKISW
TQVETGSAITWKYPSCVLVGDNSVGEFYSIALTNNKQQADTGTKMIHIGKNTRSIIISKGISAGNSANSYRGLVKMG
PKAQGARNYSQCDSMLIGDRAAANTFPYIQVDNNTAKVEHEASTSKIGEDQLFYFAQRGISEEDAVSMLVSGFCKDV
LNELPMEFAAEADKLLSLKLEGTVG

Methanobacterium thermoautotrophicum str. Delta H gi:15678031 (1997)

MLRDTLKKAEKAREKKALYGEDIDLEKFIKEEAGEHEEVTRAKEVPKEVQETLLRVGVDPEERERAGTFI
QVDQSGICTTCASESIEIMGMNVALDKYSWLKDYMWKAVAVDTDKYTATTALREAEGEMGGYFIRSKPGAREVFPLQ
ACMFIGDERVMQTAHNIVIAEENSELHIITGCATGEDVSSALHVGVSEFYLKKGARITFTMVHNWAEQVEVRPRTGI
MVGDDATYINNYILTSPVKSIQSYPTAYCTGENSRVVFQSILGGQKDSVLDMGSRVILEGRGSSAEMVSRAVSKDSS
QIYSRGHLAGRVPEVKGHLECHGLVLSDDSMIYAVPELEGSATELEMSHEAAVGKIAEEEVMYLTSRGLTEEEAASM
IVRGFLSMDITGLPPELAAETKRMLDMSLKGM

Toxoplasma gondii apicoplast gi:11496534 (1995)

MKLYKYLYNKYNNNTDLFNTVRLIGGLNINMVNKLIFKQDNFIFLYIFRLNALSILNKFKQPDWCFYELP
EFAFDDISYYSIPLNVYTNKNKYKSILSKLGLELKFSENLILDVIFDSVLLLNLTTFFLIKMGLFFLSFFQSIIFYP
YLIFSYLGSIVSNTDNFFLTINSIIFNEGSFCFVMKDLNSNINLTTYFRTHSENFAQFERTLIVLSENSKLIYFEGC
SAPMFLESQLHIAIVELFIKTKANLKYSTIQNWYRGNQLGEGGLYNFTTKRGFCMDKSFLNWIQIEIGSVITWKYPS
TYLIGNKSFSNFFSLAMLSDYQVSDTGTKMLHIGKNTKSFILSKSLSFNFSFYTYRGLVTIFKTALNSYNYTECNSL
LIGCNAFTATIPYTIINNFSAYINQEATISKLELDFLFFLLHRGLNLKSTLMILIYGYCYNICSKISFELELEVPLL
IVARAQKLFY

Odontella sinensis chloroplast gi:11467432 (1995)

MTNKSNKILNTNITKLVNQPYKYGFSTVIEKDIIEKGLNEDVICLISKKKNEPKFLLEFRLKAFKKWKEM
KCPDWAQIKFSEIDYQDIIYYSAPKVKKKLNSLDEVDPELLKTFEKLGISLTEQKRLANVAIDAVFDSVSIATTFKE
ELAECGVIFSSISEAIQEYPELIEKYLGSVVPIGDNYFSALNSAVFTDGSFCYIPKDTICPLELSTYFRINDQKSGQ
FERTLIVAEKNSQVSYLEGCTAPQYDSNQLHAAVVELVALENADIKYSTVQNWYAGNNYGEGGIYNFVTKRGLCAGS
NSKISWTQVETGSNITWKYPSCLLVGDKAKGEFYSVALTNNYQQADTGSKMIHVGKNTRSRIVSKGISAGNSKNTYR
GLVNISNKAIGARNYSQCDSLLIGNLSNANTFPFISVQNPTAKIEHEASTSKIGEEQIFYFLQRGIPIEKGVELMIS
GFCQEVFTELPLEFAAEADRLLTLKLEGSVG

Cyanophora paradoxa cyanelle gi:11467282 (1995)

MVNTQSPKNSGLENLVNQPYKYGLPLIFEIETISKGLTEETIRLISEKKNEPQFMLEFRLQAYRKWLEMS
NEPEWAHLNYPKINYQDMVYYSAPKQKKKLQSLDEVDPTLLETFEKLGIPLTEQKRLANVAVDAIFDSVSVATTFKE
ELAKEGVIFCPISEAVQKYPDLIKKYLGSVVSTSDNYFSCLNAAVFSDGSFCYIPKNVRCPLELSTYFRINNGESGQ
FERTLIVADEGSYVSYLEGCTAPQFDTNQLHAAVVELVALDNAEIKYSTVQNWYAGDENGKGGIYNFVTKRGLCAGK
NSKISWTQVETGSAITWKYPSCVLLGDNSIGEFYSVALTNRYQQADTGTKMIHIGKNTRSRIISKGISAGHSQNSYR
GLVKIGPKAVGARNYSQCDSLLIGDNSQANTFPHLQIKNPTAKVEHEASTSKIGEEQIFYFLQRGINAEEAISLIIS
GFCREVFNNLPMEFALEADKLLGLKLEGSVG

Porphyra purpurea chloroplast gi:11465652 (1995)

MVNTQNQISQTSDLDYIVNQPYKYGFTTSVESEQFPRGISREVVKLISKKKNEPEYLLNFRLKAYEKWTK
MKNPKWAHLKHPNIDFNSIIYYAVPKLKKELNSLDEVDPEILDTFNKLGISLNEQKRLSNVAVDAVFDSVSIATTFK
KELAEAGVIFCSISEAIRNYPDLIQKYLGTVVPSGDNYFAALNSAVFSDGSFCYIPPDTVCPLELSTYFRINNEESG
QFERTLIVADRGSKVSYLEGCTAPQYDTNQLHAAIVELIALDDAEIKYSTVQNWYAGNKDGKGGIYNFVTKRGLCSG
KNSKISWTQVETGSAITWKYPGCILAGDNSQGEFYSVALTNNYQEADTGTKMIHIGNNTKSKIISKGISAGKSKNSY

RGLVKIGPQSFNSRNYSQCDSLLIGQSSQANTFPYIQVQNPTAKVEHEASTSKISEDQIFYFLQRGINLEESVSLMI
SGFCKDVFNELPMEFAVEADRLLSLKLEGTVG

Guillardia theta chloroplast gi:11467607 (1989)

MSDDLSKRSLRELVSQPYKYGFHTDIENEEFPKGLDEDIIKEISKLKCEPSYMLDFRLKSYILWKKMSLP
EWACLTYLNINYQDIVYYSAPKNSTKLDSLEDVDKKILETFDKLGIPLNEQKKLANVAVDAIFDSVSVGTTFKQELS
NVGVLFCPLSEATNKYSTLVEKYLGSVVPIGDNYFAALNSAVFSEGSFCYIPPNVKCPLELSTYFRINNENSGQFER
TLIIADFNSYVSYLEGCTAPMYDKNQLHAAVVELIALENAEIRYSTVQNWYSGDTNGKGGIYNFVTKRGLCAGKSSK
ISWTQVETGSAITWKYPSCILVGEDSVGEFYSVALTNNYQQADTGTKMIHVGRGSKSRIISKGISAGYSKNTYRGQV
KININALGSINNSQCDSMLIGPYSQANTYPYIQVSNAMSRVEHEASTSKIEEEQLFYFLQRGISVEQAISLLISGFC
RDVFVKLPMEFAVEADKLLSVKLEGTVG

Cyanidium caldarium chloroplast gi:11465393 (1996)

MIDRKKSSNIQNILNKPYKYGFSTEIQSEEFPKGINEEIIRLMSHKKQEPDFILKFRLKAYQIWKKMQAP
DWGHLHHNEINFNDVLCYASPKLEQGKNKAQTISEEILATFEKLGVPIKPNNKQPKIAVDAVFDSISFGTTLQKELK
EQGIIFCSISEAIKAYPNLIKKYLGSIVPAGDNYFAALNSAVFTDGSFCYIPKNIRCPVDLSTYFRINNKEAGQFER
TLIIADENSFVNYLEGCTAPQFDTNQLHAAVVELICFKNATINYSTVQNWYAGNNKGEGGVYNFVTKRGLCQGENSK
ISWTQLETGSAITWKYPSCLLKGKRSTGEFFSVTLTNNAQEADTGTKMLHFGRQSKSLVISKGISGGVSKNTYRGLV
KISGSAIYSDNRSQCDSLLIGKGSESNTYPNLHVHNSLSKVEHEAFVSRIGEEQIFYFQQRGINIEEALNMIVSGFC
QDVCNKLPMEFALEANKLLNIKLEGSIG

Thermosynechococcus elongatus BP-1 section 2/9 gi:22294033 (2002)

MSATVQSLVNQPYKYGFVTPIETETIPKGLNEDIIRLISAKKNEPEFMLEFRLRAYRQWLKMSEPQWPRV
SYPPINYQDIVYYSAPKQKEKLKSLDEVDPVLLETFEKLGIPLSEQKRLTNVAVDAIFDSVSVATTFREELAKQGII
FCSISEALQDYPELVQKYLGSVVPIGDNFYAALNSAVFSDGSFVYVPKNTRCPMELSTYFRINNGESGQFERTLIIA
DAGSYVSYLEGCTAPMFDTNQLHAAVVELVALDNAEIKYSTVQNWYAGDENGKGGIYNFVTKRGLCLGRNSKISWTQ
VETGSAITWKYPSCVLVGDNSVGEFYSVALTNHYQQADTGTKMIHIGKNTRSRIVSKGISAGHSQNSYRGLVKIGPK
ATGARNYSQCDSMLIGDTAAANTFPYIQVQNPTAQVEHEASTSKIGEDQLFYFAQRGISAEDAVSMMISGFCRDVFN
QLPMEFAVEADRLLSLKLEGSVG

Methanobacterium thermoautotrophicum gi:2622242 (1997)

MLRDTLKKAEKAREKKALYGEDIDLEKFIKEEAGEHEEVTRAKEVPKEVQETLLRVGVDPEERERAGTFI
QVDQSGICTTCASESIEIMGMNVALDKYSWLKDYMWKAVAVDTDKYTATTALREAEGEMGGYFIRSKPGAREVFPLQ
ACMFIGDERVMQTAHNIVIAEENSELHIITGCATGEDVSSALHVGVSEFYLKKGARITFTMVHNWAEQVEVRPRTGI
MVGDDATYINNYILTSPVKSIQSYPTAYCTGENSRVVFQSILGGQKDSVLDMGSRVILEGRGSSAEMVSRAVSKDSS
QIYSRGHLAGRVPEVKGHLECHGLVLSDDSMIYAVPELEGSATELEMSHEAAVGKIAEEEVMYLTSRGLTEEEAASM
IVRGFLSMDITGLPPELAAETKRMLDMSLKGM

Nostoc sp. PCC 7120 gi:17131372 (2001)

MSATVKTLVNQPYKYGFVTDIEADTIPRGLDEDVVRLISTKKNEPEFMLEFRLRAFRQWQKMTEPTWPSV
KYPPIDYQNIIYYSAPKQKKAKLNSLDEVDPTLIETFEKLGIPLSEQKRLANVAVDAIFDSVSVATTFKEKLAKDGV
IFCSISEALQEHPELIKKYLGSVVPIADNYFAALNAAVFSDGSFVYIPKGVKCPMELSTYFRINSGDTGQFERTLIV
AEEGSYVSYLEGCTAPMYDSNQLHAAVVELVALDNAEIKYSTVQNWYAGDANGKGGIYNFVTKRGLCQGVNSKISWT
QVETGSAITWKYPSCVLVGDNSVGEFYSVALTNNMQQADTGTKMIHIGKNTRSTIISKGISAGQSSNSYRGLVKINP
TAKGARNYSQCDSMLIGDNAHANTFPYIQVQNNTGKVEHEASTSKIGEDQLFFFAQRGISSEDAISMMISGFCKDVF
NQLPMEFAVEADKLLSLKLEGSVG

Odontella sinensis complete chloroplast gi:1185127 (1995)

MTNKSNKILNTNITKLVNQPYKYGFSTVIEKDIIEKGLNEDVICLISKKKNEPKFLLEFRLKAFKKWKEM
KCPDWAQIKFSEIDYQDIIYYSAPKVKKKLNSLDEVDPELLKTFEKLGISLTEQKRLANVAIDAVFDSVSIATTFKE
ELAECGVIFSSISEAIQEYPELIEKYLGSVVPIGDNYFSALNSAVFTDGSFCYIPKDTICPLELSTYFRINDQKSGQ
FERTLIVAEKNSQVSYLEGCTAPQYDSNQLHAAVVELVALENADIKYSTVQNWYAGNNYGEGGIYNFVTKRGLCAGS
NSKISWTQVETGSNITWKYPSCLLVGDKAKGEFYSVALTNNYQQADTGSKMIHVGKNTRSRIVSKGISAGNSKNTYR
GLVNISNKAIGARNYSQCDSLLIGNLSNANTFPFISVQNPTAKIEHEASTSKIGEEQIFYFLQRGIPIEKGVELMIS
GFCQEVFTELPLEFAAEADRLLTLKLEGSVG

Skeletonema costatum chromoplast ycf24 gene, partial gi:4210403 (1999)

AVFTDGSFCYIPKDVICPLDLSTYFRINDQNSGQFERTLIIAEENSKVSYLEGCTAPQYDNNQLHAAIVE
LIALKNATIKYSTVQNWYSGDQKGQGGVYNFVTKRGLCAGDFSKISWTQVETGSSITWKYPSCVLVGDSAQGEFYSV
ALTNNYQQADTGTKMIHIGRNTRSRIVSKGISA

Cyanidium caldarium strain RK1 chloroplast gi:6466296 (1996)

MIDRKKSSNIQNILNKPYKYGFSTEIQSEEFPKGINEEIIRLMSHKKQEPDFILKFRLKAYQIWKKMQAP
DWGHLHHNEINFNDVLCYASPKLEQGKNKAQTISEEILATFEKLGVPIKPNNKQPKIAVDAVFDSISFGTTLQKELK
EQGIIFCSISEAIKAYPNLIKKYLGSIVPAGDNYFAALNSAVFTDGSFCYIPKNIRCPVDLSTYFRINNKEAGQFER
TLIIADENSFVNYLEGCTAPQFDTNQLHAAVVELICFKNATINYSTVQNWYAGNNKGEGGVYNFVTKRGLCQGENSK
ISWTQLETGSAITWKYPSCLLKGKRSTGEFFSVTLTNNAQEADTGTKMLHFGRQSKSLVISKGISGGVSKNTYRGLV
KISGSAIYSDNRSQCDSLLIGKGSESNTYPNLHVHNSLSKVEHEAFVSRIGEEQIFYFQQRGINIEEALNMIVSGFC
QDVCNKLPMEFALEANKLLNIKLEGSIG

Toxoplasma gondii chloroplast gi:5231237 (1995)

MKLYKYLYNKYNNNTDLFNTVRLIGGLNINMVNKLIFKQDNFIFLYIFRLNALSILNKFKQPDWCFYELP
EFAFDDISYYSIPLNVYTNKNKYKSILSKLGLELKFSENLILDVIFDSVLLLNLTTFFLIKMGLFFLSFFQSIIFYP
YLIFSYLGSIVSNTDNFFLTINSIIFNEGSFCFVMKDLNSNINLTTYFRTHSENFAQFERTLIVLSENSKLIYFEGC
SAPMFLESQLHIAIVELFIKTKANLKYSTIQNWYRGNQLGEGGLYNFTTKRGFCMDKSFLNWIQIEIGSVITWKYPS
TYLIGNKSFSNFFSLAMLSDYQVSDTGTKMLHIGKNTKSFILSKSLSFNFSFYTYRGLVTIFKTALNSYNYTECNSL
LIGCNAFTATIPYTIINNFSAYINQEATISKLELDFLFFLLHRGLNLKSTLMILIYGYCYNICSKISFELELEVPLL
IVARAQKLFY

Guillardia theta complete plastid genome gi:3602932 (1989)

MSDDLSKRSLRELVSQPYKYGFHTDIENEEFPKGLDEDIIKEISKLKCEPSYMLDFRLKSYILWKKMSLP
EWACLTYLNINYQDIVYYSAPKNSTKLDSLEDVDKKILETFDKLGIPLNEQKKLANVAVDAIFDSVSVGTTFKQELS
NVGVLFCPLSEATNKYSTLVEKYLGSVVPIGDNYFAALNSAVFSEGSFCYIPPNVKCPLELSTYFRINNENSGQFER
TLIIADFNSYVSYLEGCTAPMYDKNQLHAAVVELIALENAEIRYSTVQNWYSGDTNGKGGIYNFVTKRGLCAGKSSK
ISWTQVETGSAITWKYPSCILVGEDSVGEFYSVALTNNYQQADTGTKMIHVGRGSKSRIISKGISAGYSKNTYRGQV
KININALGSINNSQCDSMLIGPYSQANTYPYIQVSNAMSRVEHEASTSKIEEEQLFYFLQRGISVEQAISLLISGFC
RDVFVKLPMEFAVEADKLLSVKLEGTVG

Porphyra purpurea chloroplast gi:1276652 (1995)

MVNTQNQISQTSDLDYIVNQPYKYGFTTSVESEQFPRGISREVVKLISKKKNEPEYLLNFRLKAYEKWTK
MKNPKWAHLKHPNIDFNSIIYYAVPKLKKELNSLDEVDPEILDTFNKLGISLNEQKRLSNVAVDAVFDSVSIATTFK
KELAEAGVIFCSISEAIRNYPDLIQKYLGTVVPSGDNYFAALNSAVFSDGSFCYIPPDTVCPLELSTYFRINNEESG
QFERTLIVADRGSKVSYLEGCTAPQYDTNQLHAAIVELIALDDAEIKYSTVQNWYAGNKDGKGGIYNFVTKRGLCSG
KNSKISWTQVETGSAITWKYPGCILAGDNSQGEFYSVALTNNYQEADTGTKMIHIGNNTKSKIISKGISAGKSKNSY
RGLVKIGPQSFNSRNYSQCDSLLIGQSSQANTFPYIQVQNPTAKVEHEASTSKISEDQIFYFLQRGINLEESVSLMI
SGFCKDVFNELPMEFAVEADRLLSLKLEGTVG

E.coli genomic DNA, Kohara clone #321 gi:1742768 (1996)
MWLWRKLWGIGGTMSRNTEATDDVKTWTGGPLNYKEGFFTQLATDELAKGINEEVVRAISAKRNEPEWML
EFRLNAYRAWLEMEEPHWLKAHYDKLNYQDYSYYSAPSCGNCDDTCASEPGAVQQTGANAFLSKEVEAAFEQLGVPV
REGKEVAVDAIFDSVSVATTYREKLAEQGIIFCSFGEAIHDHPELVRKYLGTVVPGNDNFFAALNAAVASDGTFIYV
PKGVRCPMELSTYFRINAEKTGQFERTILVADEDSYVSYIEGCSAPVRDSYQLHAAVVEVIIHKNAEVKYSTVQNWF
PGDNNTGGILNFVTKRALCEGENSKMSWTQSETGSAITWKYPSCILRGDNSIGEFYSVALTSGHQQADTGTKMIHIG
KNTKSTIISKGISAGHSQNSYRGLVKIMPTATNARNFTQCDSMLIGANCGAHTFPYVECRNNSAQLEHEATTSRIGE
DQLFYCLQRGISEEDAISMIVNGFCKDVFSELPLEFAVEAQKLLAISLEHSVG

E.coli genomic DNA, Kohara clone #320 gi:1742754 (1996)
MWLWRKLWGIGGTMSRNTEATDDVKTWTGGPLNYKEGFFTQLATDELAKGINEEVVRAISAKRNEPEWML
EFRLNAYRAWLEMEEPHWLKAHYDKLNYQDYSYYSAPSCGNCDDTCASEPGAVQQTGANAFLSKEVEAAFEQLGVPV
REGKEVAVDAIFDSVSVATTYREKLAEQGIIFCSFGEAIHDHPELVRKYLGTVVPGNDNFFAALNAAVASDGTFIYV
PKGVRCPMELSTYFRINAEKTGQFERTILVADEDSYVSYIEGCSAPVRDSYQLHAAVVEVIIHKNAEVKYSTVQNWF
PGDNNTGGILNFVTKRALCEGENSKMSWTQSETGSAITWKYPSCILRGDNSIGEFYSVALTSGHQQADTGTKMIHIG
KNTKSTIISKGISAGHSQNSYRGLVKIMPTATNARNFTQCDSMLIGANCGAHTFPYVECRNNSAQLEHEATTSRIGE
DQLFYCLQRGISEEDAISMIVNGFCKDVFSELPLEFAVEAQKLLAISLEHSVG

Cyanophora paradoxa cyanelle gi:1016083 (1995)
MVNTQSPKNSGLENLVNQPYKYGLPLIFEIETISKGLTEETIRLISEKKNEPQFMLEFRLQAYRKWLEMS
NEPEWAHLNYPKINYQDMVYYSAPKQKKKLQSLDEVDPTLLETFEKLGIPLTEQKRLANVAVDAIFDSVSVATTFKE
ELAKEGVIFCPISEAVQKYPDLIKKYLGSVVSTSDNYFSCLNAAVFSDGSFCYIPKNVRCPLELSTYFRINNGESGQ
FERTLIVADEGSYVSYLEGCTAPQFDTNQLHAAVVELVALDNAEIKYSTVQNWYAGDENGKGGIYNFVTKRGLCAGK
NSKISWTQVETGSAITWKYPSCVLLGDNSIGEFYSVALTNRYQQADTGTKMIHIGKNTRSRIISKGISAGHSQNSYR
GLVKIGPKAVGARNYSQCDSLLIGDNSQANTFPHLQIKNPTAKVEHEASTSKIGEEQIFYFLQRGINAEEAISLIIS
GFCREVFNNLPMEFALEADKLLGLKLEGSVG

Clearly, a number of *ycf*24 genes and *ycf*24 gene products were known and identifiable at the time of the present invention. The specification exemplifies *ycf 24* genes which may have any of SEQ ID NO: 1 (the sequence of the malaria parasite *Plasmodium falciparum*), SEQ ID NO: 2 (the sequence of *Synechocystis* PCC6803) and SEQ ID NO: 3 (the sequence of *E.coli*). See, page 4, lines 32 -34 of the specification, where it is stated that the *ycf 24* gene product is generally "one which can be expressed from the coding region of: (a) the polynucleotide sequence of SEQ ID NO: 1, 2, or 3 ...". The specification further describes that the *ycf 24* gene of the presently claimed invention may be encoded by "polynucleotide which can selectively hybridize to the

coding region of" SEQ ID NO: 1, 2 or 3 (see page 5, lines 2-3 of the specification). An example of hybridization conditions is given at page 5, lines 6-9. Moreover, the specification describes that the *ycf 24* gene may be a malaria gene, a red algal gene, a bacterial gene or an *E.coli* gene. For example, the specification makes clear at page 4, lines 26-31 that the organism may be "*Plasmodium falciparum* [a malaria parasite] ... an alga ... a bacterium ... or *E.coli*".

The present specification therefore provides a functional and a structural description of a number of *ycf*24 genes and *ycf*24 gene products. The structural similarity of the proteins and DNA sequences of the claims at issue in this appeal are described by reference to the well known identification "*ycf*24 gene" and exemplification of a number of sequences falling within the definition.

The Patent Office's "current understanding... regarding the written description requirement of 35 U.S.C. 112, ¶1" (see, 66 FR 1099, Friday, January 5, 2001 (copy attached as Appendix E) states that

> "An applicant may show possession of an invention by disclosure of drawings[39] or structural chemical formulas[40] that are sufficiently detailed to show that applicant was in possession of the claimed invention as a whole. The description need only describe in detail that which is new or not conventional.[41] This is equally true whether the claimed invention is directed to a product or a process.
>
> An applicant may also show that an invention is complete by disclosure of sufficiently detailed, relevant identifying characteristics[42] which provide evidence that applicant was in possession of the claimed invention,[43] i.e., complete or partial structure, other physical and/or chemical properties, functional characteristics when coupled with a known or disclosed correlation between function and structure, or some combination of such characteristics.[44] What is conventional or well-known to one of ordinary skill in the art need not be disclosed in detail.[45] If a skilled artisan would have understood the inventor to be in possession of the claimed invention at the time of filing, even if every nuance of

the claims is not explicitly described in the specification, then the adequate description requirement is met.[46]" Id. at 1106.

The indicated footnotes 39-46 further support this "understanding" of the Patent Office based on Federal Circuit, CCPA and other case law as follows:

[39]*See, e.g., Vas-Cath*, 935 F.2d at 1565, 19 USPQ2d at 1118 ("drawings alone may provide a 'written description' of an invention as required by § 112"); *In re Wolfensperger*, 302 F.2d 950, 133 USPQ 537 (CCPA 1962) (the drawings of applicant's specification provided sufficient written descriptive support for the claim limitation at issue); *Autogiro Co. of America v. United States*, 384 F.2d 391, 398, 155 USPQ 697, 703 (Ct. Cl. 1967) ("In those instances where a visual representation can flesh out words, drawings may be used in thes ame manner and with the same limitations as the specification.").

[40]*See e.g., Eli Lilly*, 119 F.3d at 1568, 43 USPQ2d at 1406 ("In claims involving chemical materials, generic formulae usually indicate with specificity what the generic claims encompass. One skilled in the art can distinguish such a formula from others and can identify many of the species that the claims encompass. Accordingly, such a formula is normally an adequate description of the claimed genus.").

[41]*See Hybritech v. Monoclonal Antibodies*, 802 F.2d at 1384, 231 USPQ at 94; *Fonar Corp. v. General Electric Co.*, 107 F.3d at 1549, 41 USP!2d at 1805 (source code description not required).

[42]For example, the presence of a restriction enzyme map of a gene may be relevant to a statement that the gene has been isolated. One skilled in the art may be able to determine when the gene disclosed is the same as or different from a gene isolated by another by comparing the restriction enzyme map. In contrast, evidence that the gene could be digested with a nuclease would not normally represent a relevant characteristic since any gene would be digested with a nuclease. Similarly, isolation of an mRNA and its expression to produce the protein of interest is strong evidence of possession of an mRNA for the protein.

For some biomolecules, examples of identifying characteristics include a sequence, structure, binding affinity, binding specificity, molecular weight, and length. Although structural formulas provide a convenient method of demonstrating possession of specific molecules, other identifying characteristics or combinations of characteristics may demonstrate the requisite possession. For example, unique cleavage by particular enzymes,

isoelectric points of fragments, detailed restriction enzyme maps, a comparison of enzymatic activities, or antibody cross-reactivity may be sufficient to show possession of the claimed invention to one of skill in the art. *See Lockwood*, 107 F.3d at 1572, 41 USPQ2d at 1966 ("written description" requirement may be satisfied by using" such descriptive means as words, structures, figures, diagrams, formulas, etc., that fully set forth the claimed invention").

[43]A definition by function alone "does not suffice" to sufficiently describe a coding sequence "because it is only an indication of what the gene does, rather than what it is." *Eli Lilly*, 119 F.3 at 1568, 43 USPQ2d at 1406. *See also Fiers*, 984 F.2d at 1169-71, 25 USPQ2d at 1605-06 (discussing *Amgen Inc. v. Chugai Pharmaceutical Co.*, 927 F.2d 1200, 18 USPQ2d 1016 (Fed. Cir. 1991)).

[44]If a claim limitation invokes 35 U.S.C. 112, ¶ 6, it must be interpreted to cover the corresponding structure, materials, or acts in the specification and "equivalents thereof." See 35 U.S.C. 112, ¶ 6. See also *B. Braun Medical, Inc. v. Abbott Lab.*, 124 F.3d 1419, 1424, 43 USPQ2d 1896, 1899 (Fed. Cir. 1997). In considering whether there is 35 U.S.C. 112, ¶ 1, support for a means- (or step) plus-function claim limitation, the examiner must consider not only the original disclosure contained in the summary and detailed description of the invention portions of the specification, but also the original claims, abstract, and drawings. A means- (or step-) plus-function claim limitation is adequately described under 35 U.S.C. 112, ¶ 1, if: (1) The written description adequately links or associates adequately described particular structure, material, or acts to the function recited in a means- (or step-) plus-function claim limitation; or (2) it is clear based on the facts of the application that one skilled in the art would have know what structure, material, or acts perform the function recited in a means- (or step-) plus- function limitation. Note also: A rejection under 35 U.S.C. 112, ¶ 2, "cannot stand where there is adequate description in the specification to satisfy 35 U.S.C. 112, first paragraph, regarding means-plus-function recitations that are not, per se, challenged for being unclear." *In re Noll*, 545 F.2d 141, 149, 191 USPQ 721, 727 (CCPA 1976). See *Supplemental Examination Guidelines for Determining the Applicability* of 35 U.S.C. 112, ¶ 6, 65 FR 38510, June 21, 2000.

[45]See *Hybritech Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d at 1384, 231 USPQ at 94.

[46]*See, e.g., Vas-Cath*, 935 F.2d at 1563, 19 USPQ2d at 1116; *Martin v. Johnson*, 454 F.2d 746, 751, 172 USPQ 391, 395 (CCPA 1972) (starting "the description need not be in *ipsis verbis* (i.e., "in the same words"] to be sufficient")." Id. at 1109-1110. (Emphasis added.)

As noted above in footnote 40, the Patent Office confirms that the court in *Eli Lilly* found that claims involving generic formula usually indicate with specificity what the generic claims encompass. The court confirmed that one of ordinary skill in the art can usually distinguish such a formula from others and can identify many of the species that the claims encompass. Given these facts, the *Eli Lilly* court concluded that "such a formula is normally an adequate written description."

In the present appeal, the appellants have described, and recited in the claims and specification, a gene and gene product by reference to three nucleic acid and three amino acid sequences, a source of these material, and a well-known identifier (i.e., *ycf*24 gene) as evidenced by at least the five references of Appendix C and as exemplified by at least the sequence excerpts from the publicly available NCBI database provided in Appendix D, which allows one of ordinary skill to distinguish the generic formula of the claims from other gene product and gene sequences. One of ordinary skill can identify many species that the claims encompass. Given the conclusions of the *Eli Lilly* court, the appellants submit that the generic formula of the claims on appeal and the specification provide an adequate written description.

The issue before the *Eli Lilly* court, which was not mentioned in the footnote of the Patent Office's "written description" analysis, was whether even more generic statements, "such as 'vertebrate insulin cDNA' or 'mammalian insulin cDNA', without more," is an adequate' written description. See, 43 USPQ2d 1406 (emphasis added).

The *Eli Lilly* court found that such a generic recitation was not an adequate written description.

> "because it does not distinguish the claimed genus from others except by function. It does not specifically define any of the genes that fall within its definition. It does not specifically define any of the genes that fall within its definition. It does not define any structural features commonly possessed by members of the genus that distinguish them from others. One skilled in the art therefore cannot, as one can do with a fully described genus, visualize or recognize the identity of the members of the genus. A definition by function, as we have previously indicated, does not suffice to define the genus because it is only an indication of what the gene does, rather than what it is. *See Fiers*, 984 F.2d at 1169-71, 25 USPQ2d at 1605-06 (discussing *Amgen*)." Id.

As noted above, the claims at issues in the present appeal recite a reference sequence, i.e., the *ycf*24 gene, and the specification exemplifies a number of sequences, which allow one of ordinary skill to distinguish the gene and gene product of the claimed invention from other genes and gene products.

The appellants respectfully submit that the present specification demonstrates possession of the claimed invention by, for example, disclosure of reference sequences (i.e., SEQ ID NOs: 1, 2 and 3) and the well-known source of these sequences coupled with the functional characteristics recited in the claims. An ordinarily skilled artisan would have understood the appellants were in possession of the claimed invention at the time of filing, even if every nuance of the claimed invention is not explicitly described in the specification.

Beyond the Patent Office "understanding" of the requirements of the Section 112, first paragraph, written description, requirement, as detailed above, the Patent Office has issued Training Materials

> "designed to aid PTO's patent examiners in applying the interim written description... guidelines in a uniform and consistent manner to promote the issuance of high quality patents. The training materials will also assist patent applicants in responding to the PTO when... written description issues are raised during the

examination of a patent application." See, Press Release #00-15, USPTO, March 1, 2000 (www.uspto.gov/web/offices/com/speeches/00-15.html) (copy attached as Appendix F).

The Written Description Training Materials (http://www.uspto.gov/web/offices/pac/writtendesc.pdf) offer the following Examples 14 and 18 ("Product by Function" and "Process claim where the novelty is in the method steps"):

"Example 14: Product by Function
Specification: The specification exemplifies a protein isolated from liver that catalyzes the reaction of **A→ B.** The isolated protein was sequenced and was determined to have the sequence as set forth in SEQ ID NO: 3. The specification also contemplates but does not exemplify variants of the protein wherein the variant can have any or all of the following:
substitutions, deletions, insertions and additions. The specification indicates that procedures for making proteins with substitutions, deletions, insertions and additions is routine in the art and provides an assay for detecting the catalytic activity of the protein.
Claim:
A protein having SEQ ID NO: 3 and variants thereof that are at least 95% identical to SEQ ID NO: 3 and catalyze the reaction of A →B.
Analysis:
A review of the full content of the specification indicates that a protein having SEQ ID NO: 3 or variants having 95% identity to SEQ ID NO: 3 and having catalytic activity are essential to the operation of the claimed invention. The procedures for making variants of SEQ ID NO: 3 are conventional in the art and an assay is described which will identify other proteins having the claimed catalytic activity. Moreover, procedures for making variants of SEQ ID NO: 3 which have 95% identity to SEQ ID NO: 3 and retain its activity are conventional in the art. A review of the claim indicates that variants of SEQ ID NO: 3 include but are not limited to those variants of SEQ ID NO: 3 with substitutions, deletions, insertions and additions; but all variants must possess the specified catalytic activity and must have at least 95% identity to the SEQ ID NO: 3.
Additionally, the claim is drawn to a protein which **comprises** SEQ ID NO: 3 or a variant thereof that has 95% identity to SEQ ID NO: 3. In other words, the protein claimed may be larger than SEQ ID NO: 3 or its variant with 95% identity to SEQ ID NO: 3. It should be

noted that "having" is open language, equivalent to "comprising". The claim has two different generic embodiments, the first being a protein which comprises SEQ ID NO: 3 and the second being variants of SEQ ID NO: 3. There is a single species disclosed, that species being SEQ ID NO: 3.
A search of the prior art indicates that SEQ ID NO: 3 is novel and unobvious. There is actual reduction to practice of the single disclosed species.
The specification indicates that the genus of proteins that must be variants of SEQ ID NO: 3 does not have substantial variation since all of the variants must possess the specified catalytic activity and must have at least 95% identity to the reference sequence, SEQ ID NO: 3. The single species disclosed is representative of the genus because all members have at least 95% structural identity with the reference compound and because of the presence of an assay which applicant provided for identifying all of the at least 95% identical variants of SEQ ID NO: 3 which are capable of the specified catalytic activity. One of skill in the art would conclude that applicant was in possession of the necessary common attributes possessed by the members of the genus.
Conclusion: The disclosure meets the requirements of 35 USC §112 first paragraph as providing adequate written description for the claimed invention.

Example 18: Process claim where the novelty is in the method steps.
Specification: The specification teaches a method for producing proteins using mitochondria from the fungus *Neurospora crassa*. In the method, mitochondria are isolated from this fungus and transformed with a mitochondrial expression vector which comprises a nucleic acid encoding a protein of interest. The protein is subsequently expressed, the mitochondria is lysed, and the protein is isolated. The specification exemplifies the expression of beta-galactosidase using the claimed method using a cytochrome oxidase promoter.
Claim:
1. A method of producing a protein of interest comprising;
obtaining *Neurospora crassa* mitochondria,
transforming said mitochondria with a expression vector comprising a nucleic acid that encodes said protein of interest,
expressing said protein in said mitochondria, and
recovering said protein of interest.
Analysis:
A review of the specification reveals that *Neurospora crassa* mitochondrial gene expression is essential to the function/operation

Table I. List of genes and ORFs of the *Odontella sinensis* chloroplast genome. The gene list proceeds clockwise from the LSC boundary of the right-hand inverted repeat (IRa). Genes are listed together with their putative gene products. ORFs with a (G+C)-content less than 30% (ORFs 25, 26a, 26b, 27, 29a, 29b, 41, 44, 46) are designated (A+T)-rich and may be non-coding. Numerical positions include the first nucleotide of the translation initiation codon as well as the last nucleotide of the stop codon. Transcription initiation and termination sites of tRNAs and rRNAs were determined according to multiple alignments and secondary structure models. Genes marked "(–)" are on the complementary strand.

Gene	Gene Product	Coding Sites	
	start IRa		1
trnP	tRNA-Pro(ugg)	267	340
ORF355		478	1545
rrn16	16S ribosomal RNA	2210	3694
trnI	tRNA-Ile(gau)	3761	3835
trnA	tRNA-Ala(ugc)	3838	3910
rrn23	23S ribosomal RNA	3976	6866
rrn5	5S ribosomal RNA	6938	7063
ycf32	ORF36	7318	7428
rpl32	50S ribosomal protein L32	7563	7739
	end IRa	7725	
psaC	PSI, Fe-S polypeptide SU VII, 9 kDa	7825	8073
ycf5	ORF312 (–)	8139	9077
ORF46	(A+T)-rich	9287	9427
rps6	30S ribosomal protein S6	9482	9772
trnN	tRNA-Asn(guu)	9852	9922
thiG	involved in thiamine biosynthesis	9944	10729
clpC	Clp protease, caseinolytic-like (–)	10894	13551
trnC	tRNA-Cys(gca)	13780	13850
trnL	tRNA-Leu(uaa)	13921	14006
rps10	30S ribosomal protein S10 (–)	14207	14518
tufA	elongation factor Tu (–)	14531	15760
rps7	30S ribosomal protein S7 (–)	15831	16307
rps12	30S ribosomal protein S12 (–)	16341	16721
rpl31	50S ribosomal protein L31 (–)	16773	16991
rps9	30S ribosomal protein S9 (–)	17005	17421
rpl13	50S ribosomal protein L13 (–)	17445	17864
rpoA	RNA polymerase α-chain (–)	17893	18831
rps11	30S ribosomal protein S11 (–)	18883	19275
rps13	30S ribosomal protein S13 (–)	19320	19691
rpl36	50S ribosomal protein L36 (–)	19713	19826
secY	preprotein-translocase subunit Y (–)	19866	21143
rps5	30S ribosomal protein S5 (–)	21166	21723
rpl18	50S ribosomal protein L18 (–)	21765	22172
rpl6	50S ribosomal protein L6 (–)	22219	22758
rps8	30S ribosomal protein S8 (–)	22777	23175

rpl5	50S ribosomal protein L5 (–)	23209	23925
rpl24	50S ribosomal protein L24 (–)	23957	24190
rpl14	50S ribosomal protein L14 (–)	24191	24556
rps17	30S ribosomal protein S17 (–)	24569	24823
rpl29	50S ribosomal protein L29 (–)	24868	25137
rpl16	50S ribosomal protein L16 (–)	25144	25557
rps3	30S ribosomal protein S3 (–)	25609	26253
rpl22	50S ribosomal protein L22 (–)	26295	26642
ORF148	(–)	26691	27137
rps19	30S ribosomal protein S19 (–)	27149	27427
rpl2	50S ribosomal protein L2 (–)	27472	28299
rpl23	50S ribosomal protein L23 (–)	28324	28632
rpl4	50S ribosomal protein L4 (–)	28625	29272
rpl3	50S ribosomal protein L3 (–)	29320	29910
dnaK	Hsp70-type chaperone	30213	32057
groEL	chaperonin, 60 kDa (–)	32125	33711
trnR	tRNA-Arg(acg)	33891	33964
trnQ	tRNA-Gln(uug)	34020	34089
psbW	PSII, protein W, 13 kDa	34326	34673
ycf40	ORF73, homologous to *Porphyra* ORF71 (–)	34997	35218
trnH	tRNA-His(gug)	35294	35368
rps4	30S ribosomal protein S4	35435	35955
rps16	30S ribosomal protein S16	36238	36477
ycf35	ORF128	36607	36993
psbA	PSII, D1 reaction-center protein (–)	37339	38421
ycf44	ORF382, homologous to *Porphyra* ORF437 (–)	38642	39790
ycf46	ORF497, homologous to *Porphyra* ORF491 (–)	39799	41292
rpl34	50S ribosomal protein L34 (–)	41404	41550
secA	preprotein-translocase subunit α (–)	41583	44249
rpl27	50S ribosomal protein L27 (–)	44272	44523
rpl21	50S ribosomal protein L21 (–)	44546	44863
ycf30	former *trsE*(*rbcR* homolog) (–)	45509	46438
trnL	tRNA-Leu(uag) (–)	46484	46564
	end IRb	46634	
rpl32'	50S ribosomal protein L32' (–)	46632	46796
ycf32	ORF36 (–)	46931	47041
rrn5	5S ribosomal RNA (–)	47296	47421
rrn23	23S ribosomal RNA (–)	47493	50383
trnA	tRNA-Ala(ugc) (–)	50449	50521
trnI	tRNA-Ile(gau) (–)	50524	50598
rrn16	16S ribosomal RNA (–)	50665	52149
ORF355	(–)	52814	53881
trnP	tRNA-Pro(ugg) (–)	54019	54092
	start IRb		54358
acp	acyl carrier protein (–)	54453	54695
ycf45	ORF455, homologous to *Porphyra* ORF565 (–)	54953	56320
rpl20	50S ribosomal protein L20 (–)	56561	56905

rpl35	50S ribosomal protein L35 (–)	56915	57109
ycf42	ORF204, homologous to *Porphyra* ORF199 (–)	57112	57726
psaE	PSI, subunit IV (–)	57835	58050
ycf25	ORF 644, homologous to *E. coli* cell-division protein FtsH	58339	60273
rps14	30S ribosomal protein S14	60428	60730
psaM	PSI, protein M	60884	60976
chlI	chlorophyll biosynthesis, probable magnesium-chelatase subunit	61054	62115
ycf47	ORF74, homologous to *Porphyra* ORF71	62143	62367
trnM	tRNA-Met(cau)	62586	62659
psaD	PSI, ferredoxin-binding protein II	62744	63163
trnS	tRNA-Ser(uga)	63199	63286
petB	cytochrome b6	63341	63988
petD	cytochrome b6/f complex, subunit IV	64035	64517
trnR	tRNA-Arg(ucu) (–)	64774	64845
trnD	tRNA-Asp(guc) (–)	65093	65167
trnS	tRNA-Ser(gcu) (–)	65253	65340
trnfM	tRNA-fMet(cau) (–)	65380	65452
ycf33	ORF64	65614	65808
ORF29a	(A+T)-rich (–)	65920	66009
ORF41	(A+T)-rich (–)	66057	66182
trnY	tRNA-Tyr(gua) (–)	66350	66431
trnV	tRNA-Val(uac)	66555	66626
trnT	tRNA-Thr(ugu)	66647	66717
rbcS	Rubisco, small subunit (–)	66864	67283
rbcL	Rubisco, large subunit (–)	67323	68795
ycf24	ORF486	69086	70546
ycf16	ORF251	70546	71301
atpI	ATP synthase CF_0 subunit IV	71429	72157
atpH	ATP synthase CF_0 subunit III	72227	72475
atpG	ATP synthase CF_0 subunit II	72587	73057
atpF	ATP synthase CF_0 subunit I	73121	73660
atpD	ATP synthase CF_1 subunit δ	73657	74220
atpA	ATP synthase CF_1 subunit α	74268	75779
psbB	PSII, CP47 chlorophyll apoprotein	76048	77577
psbT	PSII, protein T, 3 kDa(*ycf8*)	77629	77727
psbN	PSII, protein N (–)	77752	77883
psbH	PSII, phosphoprotein, 10 kDa	77966	78169
ycf6	ORF29	78255	78344
ycf31	ORF42	78384	78512
psaJ	PSI, subunit IX, 5 kDa (–)	78575	78700
psaF	PSI, subunit III, plastocyanin-binding (–)	78732	79289
psbI	PSII, protein I, 4.8 kDa (–)	79654	79770
ycf41	ORF113, homologous to *Porphyra* ORF111 (–)	79855	80196
ycf39	ORF319 (–)	80212	81171
cfxQ	involved in Rubisco-expression (–)	81201	82076
psaL	PSI, subunit XI (–)	82171	82623
ycf7	ORF31	82800	82895

ycf4	ORF181	82917	83462
trnG	tRNA-Gly(ucc)	83508	83578
psbE	cytochrome b559 α-chain	83634	83888
psbF	cytochrome b559 β-chain	83901	84032
psbL	PSII, protein L	84068	84184
psbJ	PSII, protein J	84209	84328
ORF26a	(A+T)-rich (–)	84426	84506
ORF380		84538	85680
ORF25	(A+T)-rich (–)	85745	85822
trnE	tRNA-Glu(uuc) (–)	85882	85953
trnG	tRNA-Gly(gcc) (–)	85995	86066
ycf9	ORF61 (–)	86186	86371
ycf12	ORF34 (–)	86419	86523
trnK	tRNA-Lys(uuu)	86730	86801
ORF44	(A+T)-rich	86932	87066
psbC	PSII, CP43 chlorophyll apoprotein (–)	87163	88578
psbD	PSII, D2 reaction-center protein (–)	88526	89581
psaI	PSI, subunit VIII (–)	89824	89940
psbK	PSII, protein K	90146	90280
petG	cytochrome b6/f complex, subunit V	90417	90530
rpl12	50S ribosomal protein L12 (–)	90778	91161
rpl1	50S ribosomal protein L1 (–)	91234	91926
rpl11	50S ribosomal protein L11 (–)	91949	92374
ORF27	(A+T)-rich (–)	92456	92539
trnW	tRNA-Trp(cca) (–)	92660	92732
dnaB	DNA-replication helicase	92815	94182
trnF	tRNA-Phe(gaa) (–)	94184	94256
psbX	PSII, protein X, 4.1 kDa	94593	94709
ORF99		94783	95082
psbV	PSII, cytochrome c_{550} (*petK*)	95139	95630
ORF29b	(A+T)-rich (–)	95659	95748
rpl19	50S ribosomal protein L19	96389	96751
petF	ferredoxin	96895	97194
petA	apocytochrome f (–)	97278	98222
ycf43	ORF263, homologous to *Porphyra* ORF254 (–)	98256	99047
atpE	ATP synthase CF_1 subunit ε (–)	99151	99552
atpB	ATP synthase CF_1 subunit β (–)	99566	100993
ycf3	ORF179 (–)	101162	101701
rps18	30S ribosomal protein S18 (–)	101792	102010
rpl33	50S ribosomal protein L33 (–)	102014	102208
rps20	30S ribosomal protein S20	102420	102701
rpoB	RNA polymerase β-chain	102865	107004
rpoC1	RNA polymerase β′-chain	107028	109559
rpoC2	RNA polymerase β″-chain	109599	114044
ORF26b	(A+T)-rich	114058	114138
rps2	30S ribosomal protein S2	114159	114848
psaB	PSI, P700 apoprotein A2 (–)	114970	117171
psaA	PSI, P700 apoprotein A1 (–)	117351	119609

used, with a preference for TAA. Four genes start with GTG instead of ATG (*rps3, rps13, rpl23, rbcS*).

Identified genes are named according to Hallick and Bottomley (1983) and Hallick (1989). Open reading frames shared by homologous sequences of other chloroplast genomes are designated as *ycf* (Hallick and Bairoch, 1994), genes for tRNAs are labelled using the one-letter code of their respective amino acids. Open reading frames unique to *Odontella* are designated as *ORF*, followed by the number of encoded amino acid residues. Genes shared with the *Porphyra purpurea* chloroplast genome and the cyanelle genome of *Cyanophora paradoxa*, but unknown from land-plant chloroplast genomes, include *acp, atpD, atpG, dnaK, groEL, petF, psaE, psaF, psbV, psbW, psbX, rbcS, tufA, secY*, and *ycf16, -24, -30, -31, -32, -33, -35, -39*. In addition, several genes for ribosomal subunits (*rpl1, -3, -6, -11, -12, -18, -34, -35*, and *rps5, -6, -10, -13, -17, 20*) appear to be unique among plastid genomes of non-green algae. Whereas eight open reading frames (*ycf40-47*) have counterparts in the *Porphyra* chloroplast genome, none of these ORFs were found in the cyanelle genome.

A detailed description of the sequence and its phylogenetic implications will appear elsewhere.

Acknowledgment: This work was supported by the Deutsche Forschungsgemeinschaft (**Molekulare Evolution der Pflanzen, grants Ko 439/6-1 and 439/6-2**).

References

Hallick, R.B. 1989. Proposals for the naming of chloroplast genes. II. Update to the nomenclature of genes for thylakoid membrane polypeptides. Plant Mol. Biol. Reptr. 7:266-275.

Hallick, R.B. and A. Bairoch. 1994. Proposals for the naming of chloroplast genes. III. Nomenclature for open reading frames encoded in chloroplast genomes. Plant Mol. Biol. Reptr. 12:29-30.

Hallick, R.B. and W. Bottomley. 1983. Proposals for the naming of chloroplast genes. Plant Mol. Biol. Reptr. 1:38-43.

Kowallik, K.V. 1989. Molecular aspects and phylogenetic implications of plastid genomes of certain chromophytes, pp.101-124 in *The Chromophyte Algae: Problems and Perspectives* (Green J.C., B.S.C. Leadbeater, and W.L. Diver, eds.). The Systematics Association Special Volume 38. Clarendon Press, Oxford.

Kowallik, K.V. 1993. Origin and evolution of plastids from chlorophyll *a+c*-containing algae: Suggested ancestral relationship to red and green algal plastids, pp. 223-263 in *Origins of Plastids* (Lewin, R.A., ed.). Chapman and Hall, New York and London.

Reith, M., and J. Munholland. 1995. Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. Plant Mol. Biol. Reptr. 13:338-340.

Stirewalt, V.S., C. B. Michalowski, W. Löffelhardt, H. J. Bohnert, and D. A. Bryant. 1995. Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. Plant Mol. Biol. Reptr. 13:327-332.

Valentin, K., R.A. Cattolico and K. Zetsche. 1993. Phylogenetic origin of the plastids, pp. 193-212 in *Origins of Plastids* (Lewin, R.A., ed.). Chapman and Hall, New York and London.

Plastid Genomes of Three Non-Green Algae Are Sequenced

The total sequences of the chloroplast genomes of *Marchantia polymorpha* (Ohyama et al., 1986) and *Nicotiana tabacum* (Shinozaki et al., 1986) were reported just nine years ago (see box borrowed from Stirewalt et al., 1995). Apart from the excitement caused by having complete sequences of the largest genomes to that time, these historic discoveries shifted our focus from the question, *What genes are present?* to *How is their expression regulated?* The sequencing of plastid genomes of liverwort and tobacco were followed by those of rice (Hiratsuka et al., 1989), black pine (Wakasugi et al., 1994), *Epifagus virginiana* (Wolfe et al., 1992), *Euglena gracilis* (Hallick et al., 1993), *Zea mays* (Maier et al., 1995), and now *Chlorella ellipsoidea* (M. Sugiura, personal communication).

The accumulated data permit us to make detailed comparisons among plastid genomes: The similarities between vascular and non-vascular plants and between dicots and monocots are far more abundant than are the differences.

- The numbers and kinds of genes differ only slightly.
- Many of the unidentified open reading frames are conserved.
- The occurrence of gene clusters and transcription units are highly conserved.
- The plastid genome of *Epifagus virginiana* has lost all genes encoding photosynthetic components, as one might expect of a non-photosynthetic parasite, but all genes encoding rRNAs and many genes encoding tRNAs are still present and remain functional.

Considering that euglenids have been separated for a billion years from the line that produced higher plants, we expect and find the greatest

---

Note: Following in the path of *Arabidupdate*, *OrganelleIntel* is a new feature of the *PMBR*, and is devoted to new developments in the molecular biology of plant organelles. Suggestions and contributions should be directed to *pmbr@mbcl.rutgers.edu*.

differences in the chloroplast genome of *E. gracilis*. But the differences between the plastid genome of *E. gracilis* and those of higher plants are minor compared to the similarities: genes encoding ribosomal nucleic acids in the *Euglena* plastid are organized in direct rather than inverted repeats, and introns are uncharacteristically abundant in protein-encoding genes.

Totally Sequenced Plastid Genomes

Vascular plants
- *Epifagus virginiana* (a non-green plant that parasitizes the roots of beech)
- *Marchantia polymorpha* (a liverwort)
- *Nicotiana tabacum* (tobacco)
- *Oryza sativa* (rice)
- *Pinus thunbergii* (black pine)
- *Zea mays* (maize)

Algae
- *Chlorella ellipsoidea* (a green alga)
- *Cyanophora paradoxa* (an alga with an enigmatic phylogeny)
- *Euglena gracilis* (a nominally green alga with a distinct phylogeny)
- *Odontella sinensis* (a brown alga)
- *Porphyra purpurea* (a red alga)

We now have the sequences of the plastid genomes of three non-green algae: the Glaucocystophyte *Cyanophora paradoxa* (Stirewalt et al., 1995), the red alga *Porphyra purpurea* (Reith & Munholland, 1995), and the diatom *Odontella sinensis* (Kowallik et al., 1995). Gene maps and brief descriptions of each of these genomes appear in the *Genetics Resources* section of this issue. Unlike the plastids of *E. gracilis*, (which may have originated as an endosymbiotic green alga), these plastids are very, *very* different from the plastids of green algae, and their genomes are correspondingly surprising. Those of us who had hoped for evidence of independent modes of evolution from ancient symbiotic events are not disappointed. Information for protein transport, for example, is encoded within the plastid genome of *C. paradoxa;* localization of fatty acid synthesis in the plastids of higher plants is reflected in the plastid genome of *P. purpurea* by the occurrence of genes encoding enzymes of fatty acid synthesis. Additionally, the *P. purpurea* plastid contains twice as many genes as do plastids of higher plants.

To help us browse among the new gene maps, Michael Reith has constructed a table of genes and gene products (Table I) that occur in these plastid genomes but are absent from plastids of higher plants and *E. gracilis*.

The Search Continues

The story is not over. Green algae, euglenids, red algae, brown algae, and the taxonomically singular Glaucocystophyte are only five among many algal phyla that are equally dissimilar: the more we learn about algae, the more we recognize their diversity.

Dinoflagellates should be high on the list of candidates for sequencing; the lack of histones in their nuclear chromosomes already sets them apart from all other eukaryotes. The plastids of some dinoflagellates contain unique light-harvesting complexes, such as peridinin; some plastids appear to be like those of red algae, while still others are pigmented like cyanobacteria.

The plastids of certain cryptomonads and chlorarachnids occur within a larger membrane system that also contains a nucleomorph, which can contain several hundred kb of DNA (Eschbach et al., 1991) and whose genes encode 70 S and 80 S ribosomes (McFadden et al., 1994).

Chloromonads are thought to have departed from the ancestors of green algae very early; their plastid genomes could provide clues to the evolution of chlorophyll *b*-containing lines.

Within the green algal line are the Dasycladiales, which include *Acetabularia mediterranea*, a 10-cm cell that contains a single nucleus. Before the tragic death of Hans Schweiger, his group had assembled clear evidence of genes in *A. mediterranea* that were unknown in plastids of higher plants. Tymms & Schweiger (1985) reported that its plastid genome appeared "significantly larger than 400 kb." Considering that the genome of the bacterial parasite *Myxoplasma genetalia* has only 580 kb, one could imagine the plastid of *A. mediterranea* to be nearly self-sufficient.

What can we expect to learn from further investigation of plastid diversity? Will additional patterns of genes retained in plastids offer more solid clues concerning how genes were or were not lost from the hypothetical ancient symbiont(s); or will the evolution of modern plastids be seen as a totally random process? Can we reconstruct discontinued pathways within plastids? Could we engineer more efficient plastids in higher plants with genes borrowed from algae?

Acknowledgments: The simultaneous presentation of the three articles on plastid genomes of non-green algae was organized by Hans Bohnert. Stimulating discussions on algae and algal plastids with Hans Bohnert, Michael Reith, and Richard Triemer provided the basis for raising these issues. Any mistakes are ours

—Ellen M. Reardon and C. A. Price

References

Eschbach, S., C.J.B. Hofmann, U.G. Maier, P. Sitte, P. Hansmann. 1991. A eukaryotic genome of 660 kb electrophoretic karyotype of nucleomorph and cell nucleus of the cyromonad alga *Pyrenomonas salina*. Nucleic Acids Res. 19:1779-1782.

Hallick, R.B., A. Bairoch. 1994. Proposals for naming of chloroplast genes. III. Nomenclature for open reading frames encoded in chloroplast genomes. Plant Mol. Biol. Reptr. 12:2 Suppl.):S29-S30.

Hallick, R.B., L. Hong, R.G. Drager, M.R. Favreau, A. Montfort, B. Orsat, A. Spielmann, E. Stutz. 1993. Complete sequence of *Euglena gracilis* chloroplast DNA. Nucleic Acids Res 21:3537-3544.

Hiratsuka, J., H. Shimakda, R. Whittier, T. Ishibashi, M. Sakamoto, M. Mori, C. Kondo, Y. Honji, C. Sun, B.-Y. Meng, A. K. Y.-Q. Li, Y. Nishizawa, A. Hirai, K. Shinozaki, M. Sugiura. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of cereals. Mol Gen Genet 217:185-194.

Kowallik K.V., B. Stoebe, I. Schaffran, U. Freier. 1995. The chloroplast genome of a chlorophyll *a*+*c*-containing alga, *Odontella sinensis*. Plant Mol. Biol. Reptr. 13:336-342.

Löffelhardt, W. and H.J. Bohnert. 1994. Molecular biology of cyanelles. In: *The Molecular Biology of Cyanobacteria*, (ed., D.A. Bryant), pp. 65-89, Kluwer Academic Publ., Dordrecht.

Ludwig, M., S.P. Gibbs. 1989. Evidence that the nucleomorphs of *Chlorarachnion reptans* (Chlorarachniophyceae) are vestigial nuclei: Morphology, division and DNA-DAPI fluorescence. J. Phycol. 25:385-394.

Maier, R.M., K. Neckermann, G.I. Igloi and H. Kössel. 1995. Complete sequence of the maize chloroplast genome: Gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. J. Mol. Biol. 251:614-628.

McFadden, G.I., P.R. Gilson, S.E. Douglas. 1994. The photosynthetic endosymbiont in cryptomonad cells produces both chloroplast and cytoplasmic-type ribosomes. J. Cell Sci. 107:649-657.

Ohyama, K., H. Fukuzawa, T. Kohchi, H. Shirai, T. Sano, S. Sano, K. Umesono, Y. Shiki, M. Takeuchi, Z. Chang, S. Aota, H. Inokuchi, H. Ozeki. 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322:572-574.

Reith, M., and J. Munholland. 1993. A high-resolution gene map of the chloroplast genome of the red alga *Porphyra purpurea*. Plant Cell 5:465-475.

Reith, M., and J. Munholland. 1995. Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. Plant Mol. Biol. Reptr. 13:333-345.

Reith, M. 1995. Molecular biology of rhodophyte and chromophyte plastids. Annu. Rev. Plant Physiol. Plant Mol. Biol. 46:549-575.

Stirewalt, V.L., C.B. Michalowski, W. Löffelhardt, H.J. Bohnert, D.A. Bryant. 1995. Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. Plant Mol. Biol. Reptr. 13:327-332.

Tymms, M.J., H.-G. Schweiger. 1985. Tandemly repeated nonribosomal DNA sequences in the chloroplast genome of an *Acetabularia mediterranea* strain. Proc. Natl. Acad. Sci USA 82:1706-1710.

Wakasugi, T., J. Tsudzuki, S. Ito, M. Shibata, M. Sugiura. 1994. A physical map and clone bank of the black pine (*Pinus thunbergii*)chloroplast genome. Plant Mol. Biol. Reptr. 12:227-241

Wolfe, K.H., C.W. Morden and J.D. Palmer. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc. Natl. Acad. Sci. U.S.A. 89:10648-10652.

Table I. Genes and gene products in plastid genomes of non-green algae. The genes listed were identified in the plastid genomes of *Cyanophora paradoxa*, *Odentella sinensis*, or *Porphyra purpurea*. The list does not include genes that commonly occur in plastid genomes of higher plants. The gene symbols are drawn in part from bacterial gene nomenclature; *ycf* is a system of temporary gene designations assigned to conserved reading frames in chloroplast or plastid genomes (Hallick & Bairoch, 1994). Several of the gene symbols have not yet been approved by the Commission on Plant Gene Nomenclature (http://probe.nalusda.gov:8300/cgi-bin/browse/mendel). The table was prepared by Michael Reith in consultation with Hans Bohnert and Klaus Kowallik.
RF, reading frame; *C. paradoxa* = *Cyanophora paradoxa*; *O. sinensis* = *Odontella sinensis*; *P. purpurea* = *Porphyra purpurea*.

Gene	Gene Product
accA	acetyl-CoA carboxylase carboxytransferase, α subunit
accB	acetyl-CoA carboxylase biotin carboxyl carrier protein subunit
accD	acetyl-CoA carboxylase carboxytransferase β subunit
acpP	acyl carrier protein
apcA	allophycocyanin α subunit
apcB	allophycocyanin β subunit
apcD	allophycocyanin γ subunit
apcE	phycobilisome core linker polypeptide
apcF	allophycocyanin B18 subunit
argB	acetylglutamate kinase
atpD	ATP synthase CF_1 δ-subunit
atpG	ATP synthase CF_1 γ-subunit
carA	carbamoyl phosphate synthase small subunit
chlB	protochlorophyllide reductase chlB chain
chlI	magnesium chelatase subunit
chlL	protochlorophyllide reductase iron-sulfur ATP-binding protein
chlN	protochlorophyllide reductase ChlN chain
clpC	clp protease ATP-binding subunit
cpcA	phycocyanin α subunit
cpcB	phycocyanin β subunit
cpcG	phycobilisome rod-core linker polypeptide
cpeA	phycoerythrin α subunit
cpeB	phycoerythrin β subunit
crtE	geranylgeranyl pyrophosphate synthase (*Ggps1*)
dnaB	replication helicase subunit
dnaK	hsp70-type chaperone
fabH	β-ketoacyl-acyl carrier protein synthase III
ftrB	ferredoxin-thioredoxin reductase β subunit
ftsW	putative cell (organelle) division protein
glnB	nitrogen regulatory protein PII
gltB	glutamate synthase (GOGAT)
groEL	60-kDa chaperonin

groES 10-kDa chaperonin
hemA 5-aminolevulinic acid synthase
hisH histidinol-phosphate aminotransferase
hisP histidine transport ATP-binding protein
ilvB acetohydroxyacid synthase large subunit
ilvH acetohydroxyacid synthase small subunit
infB initiation factor 2
infC initiation factor 3
nadA quinolinate synthetase
odpA pyruvate dehydrogenase E_1 component, α subunit
odpB pyruvate dehydrogenase E_1 component, β subunit
pbsA heme oxygenase
petJ cytochrome c_{553}
pgmA phosphoglycerate mutase
preA prenyl transferase
psaD photosystem I, ferredoxin-binding protein, subunit II
psaE photosystem I, subunit IV, 18- to 20 kDa
psaF plastocyanin-binding protein, subunit III
psaK photosystem I, PSI-K polypeptide ('P37')
psaL photosystem I reaction center subunit XI
psbU 9- or 12-kDa protein of oxygen-evolving complex
psbV cytochrome c_{550} (oxygen-evolving complex component)
psbW photosystem II protein W (13 kDa)
psbX photosystem II protein X (4.1 kDa)
rbcS ribulose-bisphosphate carboxylase, small subunit
rne RNAse E
rnpB RNA component of RNAse P
rpl1, -3, -4, -5, -6, -9, -11, -12, -13, -18, -19, -24, -27, -28, -29, -31, -34, -35
ribosomal proteins L3–L35
rps1, -5, -6, -9, -10, -13, -17, -20
ribosomal proteins S5–S20
secA preprotein translocase subunit
secY preprotein translocase subunit
syfB phenylalanine tRNA synthetase
syh histidine tRNA synthetase
thiG thiG protein, thiamine biosynthesis
trnL(GAG) transfer RNA leu
trnR(CCU) transfer RNA arg
trnS(CGA) transfer RNA ser
trpA tryptophan synthase α subunit
trpG anthranilate synthase component II
trxA thioredoxin
tsf elongation factor Ts
tufA elongation factor Tu
ycf16 ABC transporter protein in ATPase operon
ycf17 hypothetical RF 17: ~50 aa—similar to CAB/ELIP/HLIP proteins (in *P. purpurea*, *C. paradoxa*, and *C. caldarium*)
ycf18 hypothetical RF 18: ~58 aa (in *P. purpurea* and *Aglaothamnion* sp.)
ycf19 hypothetical RF 19: ~95 aa (in *P. purpurea* and *C. caldarium*)

ycf20 hypothetical RF 20: ~100 aa (in *P. purpurea*, *Aglaothamnion* sp. and *C. caldarium*)
ycf21 hypothetical RF 21: ~175 aa (in *P. purpurea*, *C. paradoxa*, and *Antithamnion* sp.)
ycf22 hypothetical RF 22: ~200 aa (in *P. purpurea* and *Antithamnion* sp.)
ycf23 hypothetical RF 23: ~265 aa (in *P. purpurea*, *C. paradoxa* and *Antithamnion* sp.)
ycf24 hypothetical RF 24: ~487 aa (in *P. purpurea*, *C. paradoxa*, *O. sinensis*, others)
ycf25 hypothetical RF 25: *E. coli ftsH* homolog (in *P. purpurea*, *O. sinensis*)
ycf26 hypothetical RF 26: *E. coli envZ* homolog —putative histidine kinase
ycf27 hypothetical RF 27: *ompR* homolog: putative transcriptional regulatory protein
ycf28 hypothetical RF 28: ntcA homolog: putative transcriptional regulatory protein
ycf29 hypothetical RF 29: *tctD* homolog: putative transcriptional regulatory protein
ycf30 hypothetical RF 30: *rbcR* homolog:putative transcriptional regulatory protein
ycf31 hypothetical RF 31: ~32 aa (in *P. purpurea*, *C. paradoxa*)
ycf32 hypothetical RF 32: ~37 aa (in *P. purpurea*, *C. paradoxa*)
ycf33 hypothetical RF 33: ~66 aa (in *P. purpurea*, *C. paradoxa*)
ycf34 hypothetical RF 34: ~76 aa (in *P. purpurea*, *C. paradoxa*)
ycf35 hypothetical RF 35: 128 aa (in *P. purpurea*, *C. paradoxa*)
ycf36 hypothetical RF 36: ~160 aa (in *P. purpurea*, *C. paradoxa*)
ycf37 hypothetical RF 37: ~172 aa (in *P. purpurea*, *C. paradoxa*)
ycf38 hypothetical RF 38: 291 aa (in *P. purpurea*, *C. paradoxa*)
ycf39 hypothetical RF 39: ~320 aa (in *P. purpurea*, *C. paradoxa*)
ycf40 hypothetical RF 40: ~72 aa (in *P. purpurea*, *O. sinensis*)
ycf41 hypothetical RF 41: ~112 aa (in *P. purpurea*, *O. sinensis*)
ycf42 hypothetical RF 42: ~200 aa (in *P. purpurea*, *O. sinensis*)
ycf43 hypothetical RF 43: ~260 aa (in *P. purpurea*, *O. sinensis*)
ycf44 hypothetical RF 44: ~400 aa (in *P. purpurea*, *O. sinensis*)
ycf45 hypothetical RF 45: ~500 aa (in *P. purpurea*, *O. sinensis*)
ycf46 hypothetical RF 46: ~495 aa (in *P. purpurea*, *O. sinensis*)
ycf47 hypothetical RF 47: ~72 aa (in *P. purpurea*, *O. sinensis*)

J Mol Evol (1999) 48:236–244

JOURNAL OF MOLECULAR EVOLUTION

The Plastid Genome of the Cryptophyte Alga, *Guillardia theta:* Complete Sequence and Conserved Synteny Groups Confirm Its Common Ancestry with Red Algae

Susan E. Douglas, Susanne L. Penny

Institute for Marine Biosciences, 1411 Oxford Street, Halifax, Nova Scotia, Canada B3H 3Z1

Received: 16 June 1998 / Accepted: 20 August 1998

Abstract. The plastid genome of the cryptophyte alga *Guillardia theta* (121,524 bp) has been completely sequenced. The genome is 33% G+C and contains a short, nonidentical inverted repeat (4.9 kb) encoding the two rRNA cistrons. The large and small single-copy regions are 96.3 and 15.4 kb, respectively. Forty-six genes encoding proteins for photosynthesis, 5 genes for biosynthetic function, 5 genes involved in replication and division, 30 tRNA genes, 44 ribosomal protein genes (26 large subunit and 18 small subunit), 3 translation factors, 8 genes encoding components of the transcriptional machinery including 3 *ycfs* (hypothetical chloroplast genes), and 26 additional *ycfs* have been identified. There are eight ORFs larger than 50 amino acids, 3 of which have homologues on the plastid genome of the rhodophyte, *Porphyra purpurea* (Reith and Munholland 1995) and/or the *Synechocystis* genome (Kaneko et al. 1996) and can be designated new *ycfs*. Intergenic spacers are very short, no introns have been detected, and several genes overlap, all resulting in a very compact genome. In addition, large clusters of genes (such as those for the ribosomal proteins) are organized into single transcriptional units (Wang et al. 1997), again resulting in an economically organized genome. The cryptophyte plastid genome is almost completely comprised of clusters of genes that are found on the rhodophyte *Porphyra purpurea*, confirming its common ancestry with red algae. Furthermore, recombination events involving both tRNA genes and the rRNA cistrons appear to have been responsible for the structure of the cryptophyte plastid genome, including the formation of the inverted repeat.

Key words: Algae — Cryptophyte — DNA sequence — Endosymbiosis — Evolution — Genome — Plastid

Introduction

Cryptophytes are an enigmatic group of small biflagellate algae that share pigment characteristics with two distinct algal groups, the rhodophytes (phycobiliproteins) and the chromophytes (chlorophyll *c*). Like chromophytes, they harbour complex plastids surrounded by four membranes, rather than the two surrounding rhodophyte and chlorophyte plastids. However, they differ from chromophytes in possessing a small nucleus-like organelle (the nucleomorph) in the reduced space between the inner and outer plastid membrane pairs (Greenwood et al. 1997).

It has been proposed that organisms containing complex plastids arose by endosymbiosis of a photosynthetic eukaryote and a phagotrophic host with subsequent loss or reduction of eukaryotic features of the endosymbiont, such as the nucleus and cytoplasm (Gillott and Gibbs 1980). The resulting plastids would contain four membranes. Cryptophytes, by possessing vestiges of these eukaryotic features in the form of a nucleomorph and periplastidal space between the inner and the outer plastid membrane pairs, could be thought of as representing an intermediate en route to complex plastids.

Correspondence to: S.E. Douglas; *e-mail:* susan.douglas@nrc.ca

Ultrastructural data (Gibbs 1981), combined with molecular sequence data (Douglas et al. 1991; Douglas and Murphy 1994), provide strong evidence that cryptophyte algae arose by secondary endosymbiosis of a primitive eukaryotic rhodophyte. Recent phylogenetic analyses have reinforced the sister-group relationship between rhodophytes and nucleomorphs and, also, demonstrated an affiliation between cryptophyte hosts and glaucocystophytes (Bhattacharya and Medlin 1995; Van De Peer et al. 1996).

Plastid gene sequences have been utilized in phylogenetic analyses aimed at determining the relationships among eukaryotic photosynthetic lineages. However, problems encountered with substitutional bias caused by the relatively high A+T content of plastid genes (Lockhart et al. 1992), as well as varying mutational rates of different plastid genes or different sites within genes (Van De Peer et al. 1996) and the possibility of lateral gene transfers (Delwiche and Palmer 1996), have yielded conflicting results that may not reflect the true evolution of these lineages. Comparisons of gene order, on the other hand, offer a means of determining relationships among plastids that are not affected by these phenomena (Kowallik 1989, 1997; Wang et al. 1997).

The recent acquisition of complete genome sequences from the plastids of a number of green (Sugiura 1992; Hallick et al. 1993; Wakasugi et al. 1997) and nongreen photosynthetic eukaryotes (Kowallik et al. 1995; Reith and Munholland 1995; Stirewalt et al. 1995), as well as the cyanobacterium *Synechocystis* PCC6803 (Kaneko et al. 1996), allows new approaches to elucidating evolutionary relationships between algal lineages. In this context, it is of special interest to investigate the coding potential of the plastid of a chlorophyll *c*- and phycobilin-containing alga that may represent an intermediate stage in the evolution of complex plastids that have arisen by secondary endosymbiosis.

Analysis of the content of the *G. theta* plastid genome reveals strong similarities with that from the rhodophyte, *Porphyra purpurea* (Reith and Munholland 1995). Large stretches of DNA are conserved in gene order between the two plastids, although reduced in size in the cryptophyte. In many cases, tRNA genes are at the borders of the conserved stretches and adjacent to genes that have been deleted in the cryptophyte plastid, indicating that rearrangements have arisen through recombination between nonhomologous tRNA genes, as described in rice (Hiratsuka et al. 1989). In addition, recombination between the directly repeated rRNA cistons of the ancestral rhodophyte plastid appears to have resulted in the formation of the inverted repeat of the present-day cryptophyte plastid.

Materials and Methods

Guillardia theta, formerly designed *Cryptomonas* Φ (Hill and Wetherbee 1990), was cultivated as described (Douglas 1988), and plastid DNA was isolated by cesium chloride equilibrium centrifugation in the presence of Hoechst 33258 (Douglas 1988). The majority of the chloroplast genome was subcloned into pUC19 (Pharmacia) using a variety of restriction enzymes. A small portion of the genome that could not be cloned into pUC19 due to a lack of appropriate restriction enzyme sites was amplified by PCR using the high accuracy polymerase Pfu (Stratagene) and cloned into the vector pCR2.1 (Invitrogen). At least three clones of each amplification product were sequenced to reduce the possibility of PCR-generated artefacts. Template DNA was prepared using the Nucleobond AX kit (Machery Nagel) and sequencing was performed using an ABI 373A automated sequencer and the AmpliTaqFS dye terminator cycle sequencing ready reaction kit (Perkin Elmer). Specific oligonucleotide primers were used to fill gaps and complete the sequence of both strands. Sequence analysis and contig assembly was performed using Sequencher (Gene Codes, Inc.). Coding regions were identified by BLAST searches of GenBank (Gish and Gates 1993) and automated database searches were performed using MAGPIE (Gaasterland and Sensen 1996). Codon usage (based on known coding sequences including *ycfs*) was calculated using DNA Strider (Marck 1988).

Results and Discussion

Genome Organization. The circular plastid DNA of *Guillardia theta* is 121,524 bp, contains two small (approximately 4.9-kb) rRNA-containing inverted repeats, and encodes 183 genes (including the duplicated rRNA cistron genes) that are equally distributed on both strands (Fig. 1). It is the epitome of compactness (90% is coding sequence), exhibiting short A+T-rich intergenic spacers, no pseudogenes or introns, and four cases of overlapping genes. This is similar to the situation in the rhodophyte *P. purpurea* and the chromophyte *Odontella sinensis* but contrasts with green plants such as rice, where only 68% of the plastid genome is coding sequence (Hiratsuka 1989), the inverted repeats are much larger, and pseudogenes and introns are commonly found (see Sugiura 1992). All except six of the open reading frames (ORFs) have homologs on at least one other plastid genome. The ORFs without clear plastid homologues include *hlpA*, which encodes a histone-like protein (Wang and Liu 1991; Grasser et al. 1997) and appears to be unique to *G. theta*, and ORFs 53, 62, 65, 125, and 252. Other than the ribosomal RNAs, no genes for structural RNAs, such as the RNA component of RNase P (*rnpB*) that is present in *Cyanophora paradoxa* (Stirewalt et al. 1995) and *P. purpurea* (Reith and Munholland 1995), have been detected.

Like most plastids, the G+C content is low (33%), and interestingly, identical to that of *P. purpurea* (Reith and Munholland 1995). The codon usage reflects this bias, with codons ending in G and C comprising only 19% of the total. However, the highly expressed genes *psbA* and *rbcL* have a different codon bias that may be a result of selection for increased translation efficiency (Morton 1998). Alternatively, the distinct codon usage of these two genes could reflect their different origin resulting from horizontal gene transfer events. Codon usage of the five unidentified ORFs is similar to that of known coding regions, indicating that they are bona fide reading frames. The ochre termination codon TAA is used in 77% of cases, with amber and opal codons being used 15

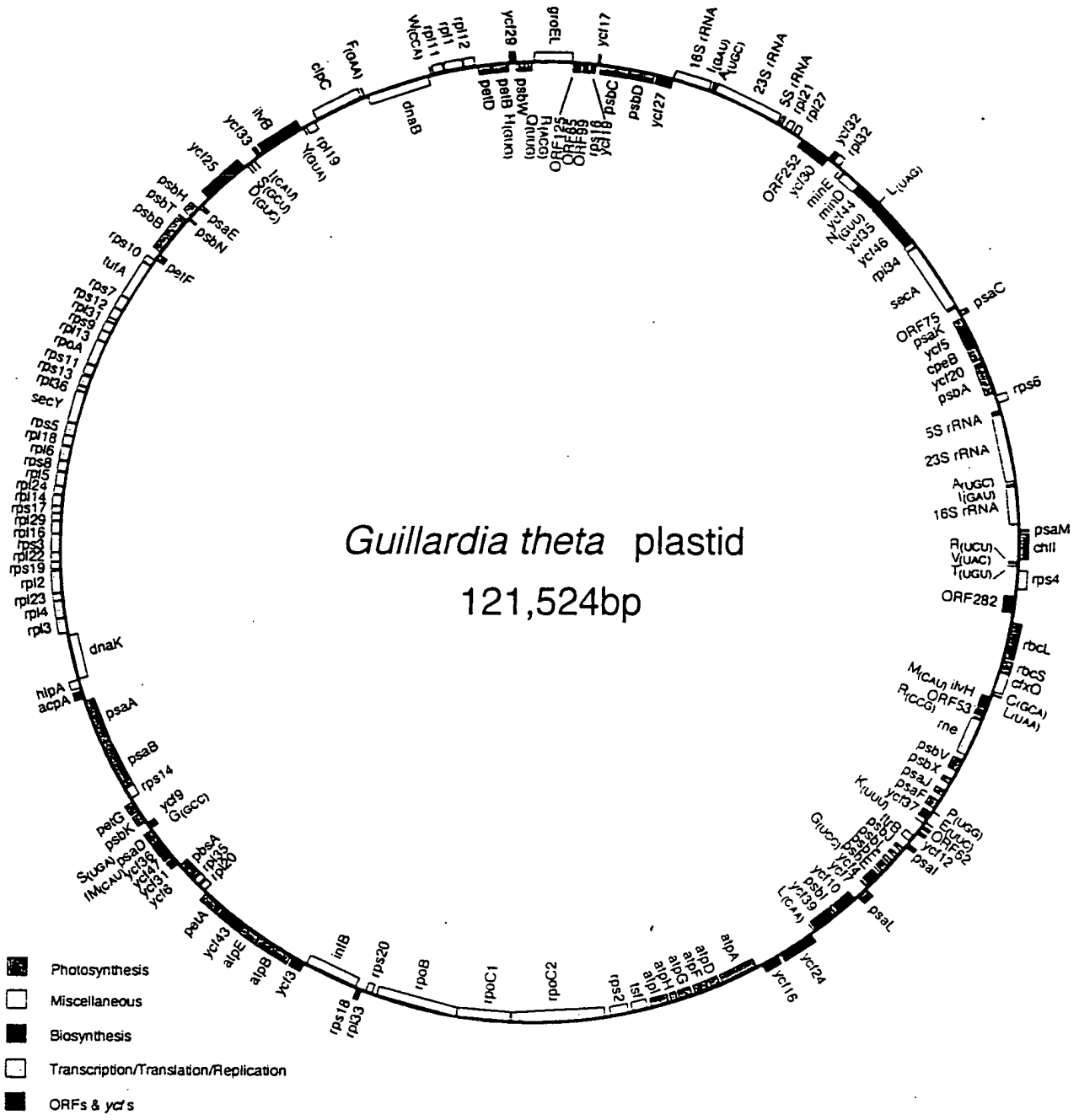


Fig. 1. Gene map of the plastid genome of *G. theta*. Genes transcribed from the plus strand are depicted outside the circle and those from the minus strand inside. Genes are shaded according to function as shown in the key.

and 8%, respectively. ATG is the predominant initiation codon, but GTG is used seven times and TTG once.

Transcription. Transcription in plastids is performed by two types of RNA polymerase, a multisubunit eubacteria-like plastic-encoded polymerase and a single-subunit phage T7-like nuclear-encoded polymerase (Allison et al. 1996). Four subunits of the eubacteria-like RNA polymerase (*rpoA, rpoB, rpoC1,* and *rpoC2*) are encoded on the *G. theta* plastid genome.

There is considerable evidence for polycistronic transcription of genes for related functions, not only from the presence of gene clusters on plastid genomes, but also from transcription studies (Douglas and Murphy 1994; Löffelhardt et al. 1997; Wang et al. 1997). Putative promoters similar to the canonical prokaryotic-type promoter sequences have been reported upstream of many *G. theta* plastid genes (Douglas et al. 1990; Douglas and Turner 1991; Douglas and Murphy 1994; Wang et al. 1997) and inspection of sequences upstream of polycistronic transcripts known from cyanobacteria and other plastids (see Löffelhardt et al. 1997) have revealed several more putative promoters. In all cases, the −10 TATAAT consensus is present, although the −35 con-

Table 1. Distribution of ycfs among photosynthetic lineages[a]

Name	Putative function	Synonym	Gr	*C.P.*	*O.s.*	*G.t.*	*P.p.*	*Syn*
ycf1	Hypothetical chloroplast RF1		+					
ycf2	Hypothetical chloroplast RF2	*ftsH* partially	+					
ycf3	Stable accumulation of PSI complex		+	+	+	+	+	+
ycf4	Stable accumulation of PSI complex		+	+	+	+	+	+
ycf5	Heme attachment to c-type cytochromes	*ccsA*	+	+	+	+	+	+
ycf6	Hypothetical chloroplast RF6		+	+	+	+	+	+
ycf7	Subunit of cytochrome b6f complex	*petI*	+	+	+	+	+	+
ycf8	PSII subunit req'd under stress	*psbT*	+	+	+	+	+	+
ycf9	Hypothetical chloroplast RF9		+	+	+	+	+	+
ycf10	Inorganic carbon uptake	*cotA/cemA*	+			+	+	+
ycf11	Acetyl-CoA carboxylase beta subunit	*aceD/zfpA*	+				+	+
ycf12	Hypothetical chloroplast RF12		+	+	+	+	+	+
ycf13	Maturase-like protein	*matA*	+					
ycf14	Hypothetical chloroplast RF14 (intron)	*matK*	+					
ycf15	Hypothetical chloroplast RF15		+					
ycf16	ABC transporter subunit			+	+	+	+	+
ycf17	Similar to CAB/ELIP/HLIP protein			+		+	+	+
ycf18	Hypothetical chloroplast RF18	*nblA*					+	+
ycf19	Hypothetical chloroplast RF19					+	+	+
ycf20	Hypothetical chloroplast RF20					+	+	+
ycf21	Hypothetical chloroplast RF21			+			+	+
ycf22	Hypothetical chloroplast RF22						+	+
ycf23	Hypothetical chloroplast RF23			+			+	+
ycf24	ABC transporter subunit			+	+	+	+	+
ycf25	Homologous to *E. coli* protein *ftsH*		+		+	+	+	+
ycf26	*envZ* homolover putative His kinase	*dfr*					+	+
ycf27	*ompR* homolover, putative trp			+	+	+	+	+
ycf28	*ntcA* homolover, putative trp						+	+
ycf29	*tctD* homolover, putative up			+		+	+	+
ycf30	*lysR* homolover, putative trp	*rbcR*		+	+	+	+	+
ycf31	Cytochrome b6f complex subunit	*petM*		+	+	+	+	+
ycf32	Photosystem II thylakoid protein			+	+	+	+	+
ycf33	Hypothetical chloroplast RF33			+	+	+	+	+
ycf34	Hypothetical chloroplast RF34			+			+	+
ycf35	Hypothetical chloroplast RF35			+	+	+	+	+
ycf36	Hypothetical chloroplast RF36			+		+	+	+
ycf37	Hypothetical chloroplast RF37			+		+	+	+
ycf38	Hypothetical chloroplast RF38			+			+	+
ycf39	Hypothetical chloroplast RF39			+	+	+	+	+
ycf40	Hypothetical chloroplast RF40				+		+	+
ycf41	Hypothetical chloroplast RF41				+		+	+
ycf42	Hypothetical chloroplast RF42	*basI*			+		+	+
ycf43	Potential integral membrane protein	*yigU/ycbT*			+	+	+	+
ycf44	c-type holocytochrome formation	*ccs*			+	+	+	+
ycf45	Hypothetical chloroplast RF45				+		+	+
ycf46	Hypothetical chloroplast RF46				+	+	+	+
ycf47	Hypothetical chloroplast RF47				+	+	+	+
ycf61	Hypothetical chloroplast RF48	ORF75				+	+	+
ycf65	Hypothetical chloroplast RF49	ORF99b				+	+	+
ycf80	Hypothetical chloroplast RF50	ORF282				+	+	

[a] Lineages or their members are abbreviated as follows: green algae and land plants, Gr; *Cyanophora paradoxa*, *C.p.*; *Odontella sinensis*, *O.s.*; *Guillardia theta*, *G.t.*; *Porphyra purpurea*, *P.p.*; and *Synechocystis* PCC, 6803, *Syn*. Presence of a *ycf* is indicated by a + symbol. trp, transcriptional regulatory protein.

sensus is sometimes absent. It is possible that a nuclear-encoded polymerase transcribes those genes where canonical prokaryotic-type promoters are absent, although the 10-nucleotide consensus promoter sequence identified from tobacco (Hajdukiewicz et al. 1997) could not be detected.

Plastid gene expression in chloroplasts is regulated mainly at the posttranscriptional level (Danon 1997). However the presence of three potential genes with significant similarity to transcriptional regulatory proteins (*ycf27*, *ycf29*, *ycf30*) in *G. theta*, *P. purpurea*, *C. paradoxa*, and *O. sinensis* plastid genomes (Table 1) and a gene for an ATP-binding polypeptide involved in the expression of Rubisco (*cfxQ*) in all except *C. paradoxa* (Table 2) indicates that at least some gene expression occurs by transcriptional regulation. Ribonuclease E

Table 2. Distribution of genes among nongreen plastid genomes[a]

Name	*C.p.*	*O.s.*	*G.t.*	*P.p.*
ATP synthase				
atpA	+	+	+	+
atpB	+	+	+	+
atpD	+	+	+	+
atpE	+	+	+	+
atpF	+	+	+	+
atpG	+	+	+	+
atpH	+	+	+	+
atpI		+	+	+
Photosystem I				
psaA	+	+	+	+
psaB	+	+	+	+
psaC	+	+	+	+
psaD		+	+	+
psaE	+	+	+	+
psaF	+	+	+	+
psaI	+	+	+	+
psaJ	+	+	+	+
psaK			+	+
psaL		+	+	+
psaM	+	+	+	+
Photosystem II				
psbA	+	+	+	+
psbB	+	+	+	+
psbC	+	+	+	+
psbD	+	+	+	+
psbE	+	+	+	+
psbF	+	+	+	+
psbH	+	+	+	+
psbI	+	+	+	+
psbJ	+	+	+	+
psbK	+	+	+	+
psbL	+	+	+	+
psbN	+	+	+	+
psbT(ycf8)	+	+	+	+
psbV	+	+	+	+
psbW	+	+	+	+
psbX	+	+	+	+
Rubisco				
rbcL	+[b]	+	+	+
rbcS	+[b]	+	+	+
Phycobiliproteins				
apcA	+			+
apcB	+			+
apcD	+			+
apcE	+			+
apcF	+			+
cpcA	+			+
cpcB	+			+
cpcG	+			+
cpeA				+
cpeB			+	+
Electron transfer				
petA	+	+	+	+
petB	+	+	+	+
petD	+	+	+	+
petF	+	+	+	+
petG	+	+	+	+
petL (ycf7)	+	+	+	+
petM (ycf31)	+	+	+	+
ftrB			+	+
Miscellaneous				
clpC		+	+	+
dnaB		+	+	+

Table 2. Continued

Name	*C.p.*	*O.s.*	*G.t.*	*P.p.*
dnaK	+	+	+	+
groEL	+	+	+	+
secA		+	+	+
secY	+	+	+	+
Ribosomal proteins				
rpl1	+	+	+	+
rpl2	+	+	+	+
rpl3	+	+	+	+
rpl4		+	+	+
rpl5	+	+	+	+
rpl6	+	+	+	+
rpl9				+
rpl11	+	+	+	+
rpl12	+	+	+	+
rpl13		+	+	+
rpl14	+	+	+	+
rpl16	+	+	+	+
rpl18	+	+	+	+
rpl19	+	+	+	+
rpl20	+	+	+	+
rpl21	+	+	+	+
rpl22	+	+	+	+
rpl23		+	+	+
rpl24		+	+	+
rpl27		+	+	+
rpl28	+			+
rpl29		+	+	+
rpl31		+	+	+
rpl32		+	+	+
rpl33	+	+	+	+
rpl34	+	+	+	+
rpl35	+	+	+	+
rpl36	+	+	+	+
rps1				+
rps2	+	+	+	+
rps3	+	+	+	+
rps4	+	+	+	+
rps5	+	+	+	+
rps6	+	+	+	+
rps7	+	+	+	+
rps8	+	+	+	+
rps9	+	+	+	+
rps10	+	+	+	+
rps11	+	+	+	+
rps12	+	+	+	+
rps13	+	+	+	+
rps14	+	+	+	+
rps16	+	+	+	+
rps17	+	+	+	+
rps18	+	+	+	+
rps19	+	+	+	+
rps20	+	+	+	+
Transcription/RNA processing				
cfxQ		+	+	+
rne			+	+
rnpB	+			+
rpoA	+	+	+	+
rpoB	+	+	+	+
rpoC1	+	+	+	+
rpoC2	+	+	+	+
Translation				
infB			+	+

Table 2. Continued

Name	C.p.	O.s.	G.t.	P.p.
tsf			+	+
tufA	+	+	+	+
Biosynthesis				
acpA	+	+	+	+
chlB	+			+
chlI	+	+	+	+
chlL	+			+
ChlN	+			+
ilvB			+	+
ilvH			+	+
pbsA			+	+
preA	+			+
trpG	+			+

[a] Except for ribosomal proteins, genes unique to a single genome are not shown. Abbreviations are as in Table 1.

[b] The rubisco genes of the *C. paradoxa* plastid are not homologous to those of the other plastids.

(*rne*), also encoded on the plastid genome of *G. theta*, may participate in posttranscriptional degradation of mRNAs.

Translation. Many components of the translational apparatus are present, including the 3 rRNA molecules, 1 initiation factor (*infB*), two elongation factors (*tsf* and *tufA*), 26 genes for 50S ribosomal subunit proteins, and 18 genes for 30S ribosomal subunit proteins. Most of these ribosomal protein genes are found in a large cluster which is conserved to different degrees in different photosynthetic lineages and has been a found to be a useful character for phylogenetic reconstruction (Sugita et al. 1997; Wang et al. 1997). In addition, several other highly conserved ribosomal protein gene clusters are present (*rpl11/1/12*, *rpl33/rps18*, *rpl20/35*, and *rpl21/27*). Thirty tRNAs are present, two of which (*trnI* and *trnA*) are duplicated in the inverted repeats. This suite of tRNAs allows the decoding of all 61 sense codons.

Photosynthesis. All of the components of the ATP synthase with the exception of *atpC*, which was transferred to the nucleus very early in the evolution of plastids (Pancic et al. 1992; Kowallik 1997), are present on the *G. theta* plastid genome. As in both *P. purpurea* and *O. sinensis*, there is an overlap of four nucleotides between the *atpF* and the *atpD* genes. Interestingly, these two genes overlap by a single nucleotide in the cyanobacterium *Synechococcus* PCC 6301 (Cozens and Walker 1987).

Seven components of the electron transfer chain are also found (*petA*, *B*, *D*, *F*, and *G*), including the recently identified *petL* [formerly designated *ycf7* (Naithani et al. 1997)] (Table 1) and *petM* [formerly designated *ycf31* (de Vitry et al. 1996)] (Table 1). With the exception of *psbM*, which has been identified only on the *C. paradoxa* plastid genome, the complete suite of 28 photosystem I and II genes is present on the *G. theta* plastid genome. Also present is the beta subunit of ferredoxin thioreductase (*ftrB*), which participates in electron transfer and regulates several photosynthetic enzymes. In addition, genes for both subunits of Rubisco (*rbcL* and *rbcS*), the beta subunit of phycoerythrin (*cpeB*), and two genes involved in chlorophyll biosynthesis—a magnesium chelatase (*chlI*) and heme oxygenase (*pbsA*)—are found on the *G. theta* plastid genome. *ycf17*, which encodes a protein similar to members of the CAB/ELIP/HLIP family, is also present.

Replication and Cell Division. Although plastids lack histones, there is evidence for chromatin-associated proteins (see Grasser et al. 1997). One such protein, encoded by *hlpA* (Wang and Liu 1991), is thought to perform an architectural role in the plastid nucleoid (Grasser et al. 1997). In addition, *dnaB*, encoding a DNA helicase, and three other genes that participate in cell division (*minD*, *minE*, *ftsH*) that were recently reported from the plastid genome of the green alga *Chlorella vulgaris* C-27 (Wakasugi et al. 1997) have been identified on the *G. theta* plastid genome. It is interesting that *minD* and *minE* have not been identified in any other nongreen algae and *hlpA* is unique among all sequenced plastids. This may indicate that the cryptophyte endosymbiont represents a primitive stage in the evolution of the nongreen lineages, just as *C. vulgaris* represents a primitive stage in the green lineage.

Miscellaneous Functions. A number of genes involved in protein metabolism or transport are encoded on the plastid genome of *G. theta*. These include *secA* and *secY*, which are components of the sec protein translocation system, *groEL* and *dnaK* (chaperonin subunits), and *clpC* (the ATP binding subunit of the Clp protease).

Conserved Reading Frames (ycfs). Of the 47 reading frames that are conserved between at least two plastid genomes and are designated by the Commission on Plant Gene Nomenclature (Hallick and Bairoch 1994) as *ycfs* (hypothetical chloroplast frames), 29 have been identified on the plastid of *G. theta*. The distribution of these among various plastid groups and the functions of those that are known are listed in Table 1. In addition, *G. theta* ORFs 76, 99, and 282 are homologous to ORFs 75, 99b, and 450 of *P. purpurea* and are now designated *ycfs* 61, 65, and 80, respectively (Stoebe, personal communication).

A Conserved Intein. The *dnaB* genes from both *G. theta* and *P. purpurea* contain an additional stretch of protein-encoding sequence that is spliced out of the mature polypeptide, much as an intron is spliced out of mature mRNA. Inteins in the *dnaB* gene are relatively rare, being found only in two eubacteria, *Synechocystis* PCC6803 and *Rhodothermus marinus* (Liu and Hu 1997). The only other inteins known to occur in plastids are in the *clpP* genes of *Chlamydomonas reinhardtii* and *C. eugametos* (Huang et al. 1994). The *G. theta dnaB* intein is 160 amino acids long, whereas the *P. purpurea*

intein is 150 amino acids long. In both organisms, the codon usage of the intein is very similar to that of the exteins, indicating that it is quite ancient and the codon usage has become homogenized over time.

The Inverted Repeat. Examination of the regions flanking the cryptophyte *rrnB* cistron shows that it contains genes from the upstream region of *rrnB* (*trnV, trnR, chlI,* and *psaM*) and the downstream region of *rrnA* (*rps6*) of *P. purpurea* (Fig. 2). The reciprocal arrangement (involving 36.9 kb of sequence upstream of the *P. purpurea rrnA* cistron) is evident in the crytophyte *rrnA* cistron. It is highly likely that the inverted repeat structure of the cryptophyte plastid has resulted from a reciprocal recombination event within the rRNA cistron.

The inverted repeats of *G. theta* are not identical in sequence. There is one transition substitution in the SSU rRNA gene, two in the $tRNA^{Ala}$–LSU rRNA intergenic-spacer, three in the LSU rRNA gene, one in the LSU rRNA–5S rRNA intergenic spacer, and two in the 5S rRNA gene. None of the substitutions affect secondary structure of the mature rRNA molecules. The regions upstream of the 16S rRNA genes are well conserved (86% similar) for 111 bp, presumably to preserve the region surrounding the promoter. However sequence similarity stops immediately downstream of the 5S rRNA gene. In *P. purpurea,* the coding sequences are much more variable than in *G. theta,* with 41 of 4820 positions differing (Reith and Munholland 1995), and the flanking sequences diverge within seven nucleotides of the 16S rRNA and within two of the 5S rRNA (Reith and Munholland 1993). This implies that the copy-correction mechanism that ensures identity of repeats in land plant chloroplasts, but is apparently absent in *P. purpurea,* may be only partially developed in *G. theta.* Similarly, the expansion of the inverted repeat by gene conversion, which has occurred to differing extents in other lineages, may have occurred to a limited extent in *G. theta* since the region upstream of the SSU rRNA is conserved for a short distance.

Synteny Groups. The entire plastid genome of *G. theta* is comprised of synteny groups that are present in *P. purpurea* (Fig. 3; junctions marked by arrowheads). Three of these synteny groups are very large (two are over 30 kb and one is 17 kb). In all cases, the gene order and transcriptional orientation are conserved, but some genes present on the plastid genome of *P. purpurea* (usually those involved in biosynthesis, phycobiliprotein synthesis or ORFs of unknown function) have been deleted from the plastid genome of *G. theta* (Fig. 4). In fact there are only five genes involved in biosynthesis remaining on the *G. theta* plastid genome (*ilvB, ilvH, pbsA, chlI,* and *acpA*), a single subunit of phycoerythrin (*cpeB*), and five ORFs of unknown function.

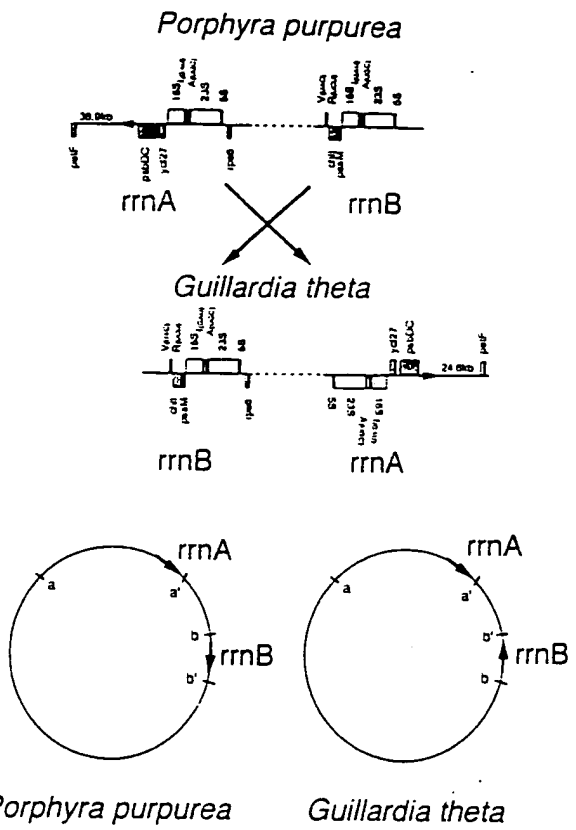


Fig. 2. Recombination between ribosomal RNA cistrons. A reciprocal crossover event between *rrn*A and *rrn*B of *P. purpurea* results in the exchange of flanking sequences and the resulting arrangement seen in *G. theta.* Deletion of several genes has resulted in the reduction in size of the *rrn*A flanking region from 36.9 kb in *P. purpurea* to 24.6 kb in *G. theta.* Borders of *rrn*A and *rrn*B are represented by a/a' and b/b', respectively.

In many cases, tRNA genes are present at the junctions of the synteny groups (asterisks; Fig. 3), suggesting that they may have participated in the deletion of gene sequences, possibly by acting as recognition signals (Hiratsuka et al. 1989). In addition, there are sixteen instances where tRNA genes are also found adjacent to *P. purpurea* genes that have been deleted from *G. theta.* Figure 4B shows one such region of the *P. purpurea* genome where three deletions have occurred relative to *G. theta,* all of which are adjacent to tRNA genes.

Evolutionary Implications. This is the first plastid genome to be sequenced form a nucleomorph-containing organism and as such it is of interest to compare its coding capacity with that of other algae that have arisen by secondary endosymbiosis but do not contain a nucleomorph. The diatom *O. sinensis* is one such example that has been completely sequenced (Kowallik et al. 1995). Although some synteny groups are shared between *G. theta* and *O. sinensis* (large ribosomal protein, *atpA, rpoBC1C2, psbBTNH* gene clusters), none are as large or as striking as those shared between *G. theta* and *P. purpurea.* There has been much more rearrangement (Fig. 4), indicating either that a longer period of evolution has

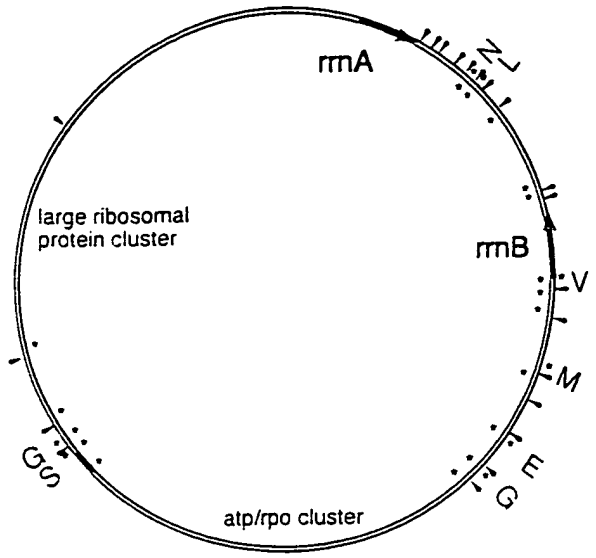


Fig. 3. Synteny groups present on the *G. theta* plastid genome. Junctions between conserved gene clusters are indicated by arrows. Asterisks inside the circle represent tRNA genes present at the junctions on the *P. purpurea* plastid genome and those outside the circle (with letters) represent those at the junctions on the *G. theta* plastid genome. The two rRNA cistrons are represented by shaded arrows and an internal inversion of the cluster *ycf6*/31/47/36/*trn*fMet within the larger synteny group is indicated by shading.

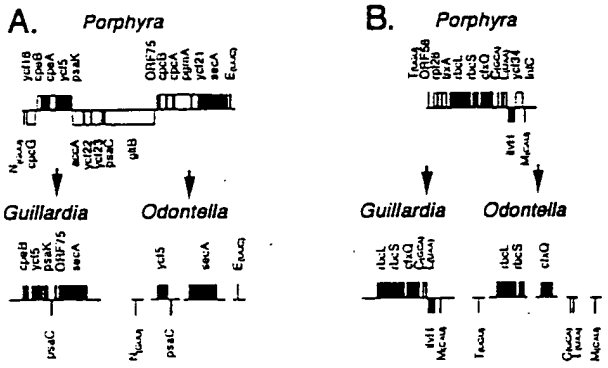


Fig. 4. (A) Gene arrangement between $tRNA^N$ and $tRNA^E$ genes of *P. purpurea*, *O. sinensis*, and *G. theta*. (B) Gene arrangement between $tRNA^T$ and $tRNA^M$ genes of *P. purpurea*, *O. sinensis*, and *G. theta*. Genes present in both *P. purpurea* and *G. theta* genomes are represented by shaded boxes and those that have been deleted from *G. theta* by empty boxes. Additional genes have been deleted from *O. sinensis*. Genes transcribed from the plus strand are depicted above the line and those from the minus strand below.

passed since the *O. sinensis* endosymbiont was established than for the *G. theta* endosymbiont or that the *O. sinensis* and *G. theta* endosymbionts were different. Alternatively, the presence of genes essential for plastid function in the nucleomorph have stabilized the plastid genome such that rearrangements do not occur to as great an extent in nucleomorph-containing organisms. That possibility is now under investigation although preliminary results indicate that there are very few genes for plastid-localized products in the nucleomorph (McFadden et al. 1997).

With the availability of complete plastid genome sequences, analysis of the distribution of gene clusters has increasingly been used for phylogenetic reconstruction (Kowallik 1997). Of particular interest are clusters that are widely separated on cyanobacterial genomes but appear to have fused subsequent to endosymbiosis and are present in plastid genomes from several lineages. Such arrangements provide very strong evidence for the monophyletic origin of plastids. For such similar organization in the plastid genomes of separate lineages to result from different endosymbionts, an extreme degree of convergent evolution would have to be invoked (Douglas 1994). Well-studied examples that have helped elucidate phylogenetic relationships between algal lineages include the large ribosomal protein cluster (Wang et al. 1997), the rRNA cistrons (Reith and Munholland 1993), the *rpoBC/atpA* cluster (Pancic et al. 1992; Kowallik 1997), and the *psbBTNH* cluster (Douglas 1994).

Three significant features suggest that the ancestor of the *G. theta* plastid closely resembled a rhodophyte plastid like that of *P. purpurea*. First, the conserved synteny groups, which are identical in gene order but reduced in gene content to stretches of the *P. purpurea* plastid genome, give strong evidence for a common ancestry. Second, both *G. theta* and *P. purpurea* contain an intein in their plastid *dnaB* genes. Given the rarity of inteins in plastid genes in general, and *dnaB* genes in particular, this is a significant shared character. Third, the inverted repeat of *G. theta* appears to have arisen by reciprocal recombination from the nonidentical, directly repeated rRNA cistrons of *P. purpurea* (interpreted by the authors as being a primitive feature) (Reith and Munholland 1993). Our results greatly strengthen their suggestions that the ancestral plastid may have had two direct, nonidentical rRNA repeats that were either reorganized into the inverted pattern seen in many land plants, glaucophytes, rhodophytes, cryptophytes, and chromophytes or reduced to a single copy in some chlorophytes and rhodophytes.

Acknowledgments. The sequence reported in this paper has been deposited in the GenBank database (accession No. AF041468) and is also available as a Magpie project at http://niji.imb.nrc.ca/magpie/plastid. The authors thank Michael Reith and Mark Ragan for comments on this manuscript and Gertraud Buger (Organelle Genome Megasequencing Project, University of Montreal) for assistance in annotation of the sequence for submission to Genbank. We are very grateful for the help of Paul Gordon, Christoph Sensen and Terry Gaasterland in the implementation of Magpie. This is NRCC publication No. 39789.

References

Allison LA, Simon LD, Maliga P (1996) Deletion of *rpoB* reveals a second distinct transcription system in plastids of higher plants. EMBO J 15:2802–2809

Bhattacharya D, Medlin L (1995) The phylogeny of plastids: A review based on comparisons of small-subunit ribosomal RNA coding regions. J Phycol 31:489–498

Cozens AL, Walker JE (1987) The organization and sequence of the genes for ATP synthase subunits in the cyanobacterium *Synechococcus* 6301. J Mol Biol 294:359–383

Danon A (1997) Translational regulation in the chloroplast. Plant Physiol 115:1293–1298

Delwiche CF, Palmer JD (1996) Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. Mol Biol Evol 13:873–882

de Vitry C, Breyton C, Pierre Y, Popot JL (1996) The 4-kDa nuclear-encoded PetM polypeptide of the chloroplast cytochrome b6f complex. Nucleic acid and protein sequences, targeting signals, transmembrane topology. J Biol Chem 271:10667–10671

Douglas SE (1988) Physical mapping of the plastid genome from the chlorophyll c-containing alga. *Cryptomonas* Φ. Curr Gen 14:591–598

Douglas SE (1994) Chloroplast origins and evolution. In: Bryant DA (ed) The molecular biology of cyanobacteria. Kluwer, Amsterdam, pp 91–118

Douglas SE, Durnford DG, Morden CW (1990) Nucleotide sequence of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from *Cryptomonas* Φ: Evidence supporting the polyphyletic origin of plastids. J Phycol 26:500–508

Douglas SE, Murphy CA (1994) Structural, transcriptional, and phylogenetic analyses of the *atpB* gene cluster from the plastid of *Cryptomonas* Φ (Cryptophyceae). J Phycol 30:329–340

Douglas SE, Murphy CA, Spencer DF, Gray MW (1991) Molecular evidence that cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. Nature 350:148–151

Douglas SE, Turner S (1991) Molecular evidence for the origin of plastids from a cyanobacterium-like ancestor. J Mol Evol 33:267–273

Gaasterland T, Sensen C (1996) Fully automated genome analysis that reflects user needs and preferences: A detailed introduction to the MAGPIE system architecture. Biochimie 78:302–310

Gibbs SP (1981) The chloroplast endoplasmic reticulum: structure, function, and evolutionary significance. Int Rev Cytol 72:49–99

Gillott MA, Gibbs SP (1980) The cryptomonad nucleomorph: Its ultrastructure and evolutionary significance. J Phycol 16:558–568

Gish W, Gates DJ (1993) Identification of protein coding regions by database similarity search. Nature Genet 3:226–272

Grasser KD, et al. (1997) The recombinant product of the *Cryptomonas* Φ plastid gene, *hlpA* is an architectural HU-like protein that promotes the assembly of complex nucleoprotein structures. Eur J Biochem 249:70–76

Greenwood AD, Griffiths HB, Santore UJ (1977) Chloroplasts and cell compartments in Cryptophyceae. Brit Phycol J 12:119

Hajdukiewicz PTJ, Allison LA, Maliga P (1997) The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. EMBO J 16: 4041–4048

Hallick RB, Bairoch A (1994) Proposals for naming of chloroplast genes. III. Nomenclature for open reading frames encoded in chloroplast genomes. Plant Mol Biol Reptr 12:S29–S30

Hallick RB, et al. (1993) Complete sequence of *Euglena gracilis* chloroplast DNA. Nucleic Acids Res 21:3537–3544

Hill DRA, Wetherbee R (1990) *Guillardia theta* new-genus new-species Cryptophyceae. Can J Bot 68:1873–1876

Hiratsuka J, et al. (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of cereals. Mol Gen Genet 27:185–194

Huang C, et al. (1994) The *Chlamydomonas* chloroplast *clpP* gene contains translated large insertion sequences and is essential for cell growth. Mol Gen Genet 244:151–159

Kaneko T, et al. (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res 3:109–136.

Kowallik KV (1989) Molecular aspects and phylogenetic implications of plastid genomes of certain chromophytes. In: Green JC et al. (ed) The chromophyte algae—problems and perspectives. Clarendon Press, Oxford, pp 101–124

Kowallik KV (1997) Origin and evolution of chloroplasts: Current status and future perspectives. In: Schenk HEA, et al. (eds) Eukaryotism and symbiosis. Springer Verlag, Berlin, Heidelberg, pp 3–23

Kowallik KV, Stoebe B, Schaffran I, Freier U (1995) The chloroplast genome of chlorophyll a+c-containing alga, *Odontella sinensis*. Plant Mol Biol Rep 13:336–342

Liu XQ, Hu Z (1997) A DnaB intein in *Rhodothermus marinus*: Indication of recent intein homing across remotely related organisms. Proc Natl Acad Sci USA 94:7851–7856

Lockhart PJ, Howe CJ, Bryant DA, Beanland TJ, Larkum AWD (1992) Substitutional bias confounds inference of cyanelle origins from sequence data. J Mol Evol 34:153–162

Löffelhardt W, Bohnert HJ, Bryant DA (1997) The complete sequence of the *Cyanophora paradoxa* cyanelle genome (*Glaucocystophyceae*). In: Bhattacharya D (ed). Springer, Wien, New York, pp 149–162

Marck C (1988) "DNA Strider": A "C" program for the fast analysis of DNA and protein sequences on Apple Macintosh family of computers. Nucleic Acids Res 16:1829–1835

McFadden GI, et al. (1997) Bonsai genomics: Sequencing the smallest eukaryotic genome. Trends Genet 13:46–49

Morton BR (1998) Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages. J Mol Evol 46:449–459

Naithani S, Trivedi PK, Sane PV (1997) Characterization of the *orf31-petG* gene cluster from the plastid genome of *Populus deltoides*. Biochem Mol Biol Int 43:433–442

Pancic PG, Strotmann H, Kowallik KV (1992) Chloroplast ATPase genes in the diatom *Odontella sinensis* reflect cyanobacterial characters in structure and arrangement. J Mol Biol 224:529–536

Reith ME, Munholland J (1993) The ribosomal RNA repeats are non-identical and directly oriented in the chloroplast genome of the red alga *Porphyra purpurea*. Curr Genet 24:443–450

Reith ME, Munholland J (1995) Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. Plant Mol Biol Rep 13: 333–335

Stirewalt VL, Michalowski CB, Loffelhardt W, Bohnert HJ, Bryant DA (1995) Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. Plant Mol Biol Rptr 13:327–332

Sugita M, et al. (1997) Organization of a large gene cluster encoding ribosomal proteins in the cyanobacterium *Synechococcus* sp. strain PCC 6301: Comparison of gene clusters among cyanobacteria, eubacteria and chloroplast genomes. Gene 195:73–79

Sugiura M (1992) The chloroplast genome. Plant Mol Biol 19:149–168

Van De Peer Y, Rensing SA, Maier UG, De Wachter R (1996) Substitution rate calibration of small subunit ribosomal RNA identifies chlorarachniophyte endosymbionts as remnants of green algae. Proc Natl Acad Sci USA 93:7732–7736

Wakasugi T, et al. (1997) Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: The existence of genes possibly involved in chloroplast division. Proc Natl Acad Sci USA 94:5967–5972

Wang SL, Liu X-Q (1991) The plastid genome of *Cryptomonas* Φ encodes an hsp70-like protein, a histone-like protein, and an acyl carrier protein. Proc Natl Acad Sci USA 88:10783–10787

Wang SL, Liu X-Q, Douglas SE (1997) The large ribosomal gene cluster of a cryptomonad plastid: Gene organization, sequence and evolutionary implications. Biochem Mol Biol Int 41:1035–1044

Plant Molecular Biology Reporter
13 (4) 1995

pages 327–332

Genetic Resource

Nucleotide Sequence of the Cyanelle Genome from *Cyanophora paradoxa*

Veronica L. Stirewalt, Christine B. Michalowski, Wolfgang Löffelhardt, Hans J. Bohnert, and Donald A. Bryant

E-mail: dab14@psuvm.psu.edu; bohnert@biosci.arizona.edu
(VLS, DAB) Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, 16802, U.S.A.
(CBM, HJB) Departments of Biochemistry, Molecular and Cellular Biology, and Plant Sciences, The University of Arizona, Tucson, AZ 85721, U.S.A.
(WL) Institut für Biochemie und Molekulare Zellbiologie und Ludwig-Boltzmann-Forschungsstelle für Biochemie, Universität Wien, A-1030, Vienna, Austria.

Key Words: *Cyanophora paradoxa*, cyanelle, chloroplast, plastid, nucleotide sequence.

Abstract: The complete nucleotide sequence of the cyanelle genome of *Cyanophora paradoxa* Pringsheim strain LB 555 was determined (accession number U30821). The circular molecule is 135,599 base pairs in length. The physical map of this DNA molecule is shown along with identified genes and open reading frames.

Cyanophora paradoxa (*Glaucocystophyceae*) is a bi-flagellated protist that contains cyanobacterium-like plastids termed cyanelles. The cyanelles of *Cyanophora paradoxa* are conspicuous because they are surrounded by a lysozyme-sensitive peptidoglycan wall that is typical of those associated with cyanobacteria but that is not found in other plastid types. Thus, this organism has frequently been considered a "living fossil" and a paradigm for the invasion of a eukaryotic cell by a cyanobacterium (for a review, see Löffelhardt and Bohnert, 1994). The complete nucleotide sequence, 135,599 bp, of the cyanelle DNA from the

Abbreviations: LSC, large single-copy region; SSC, small single-copy region.

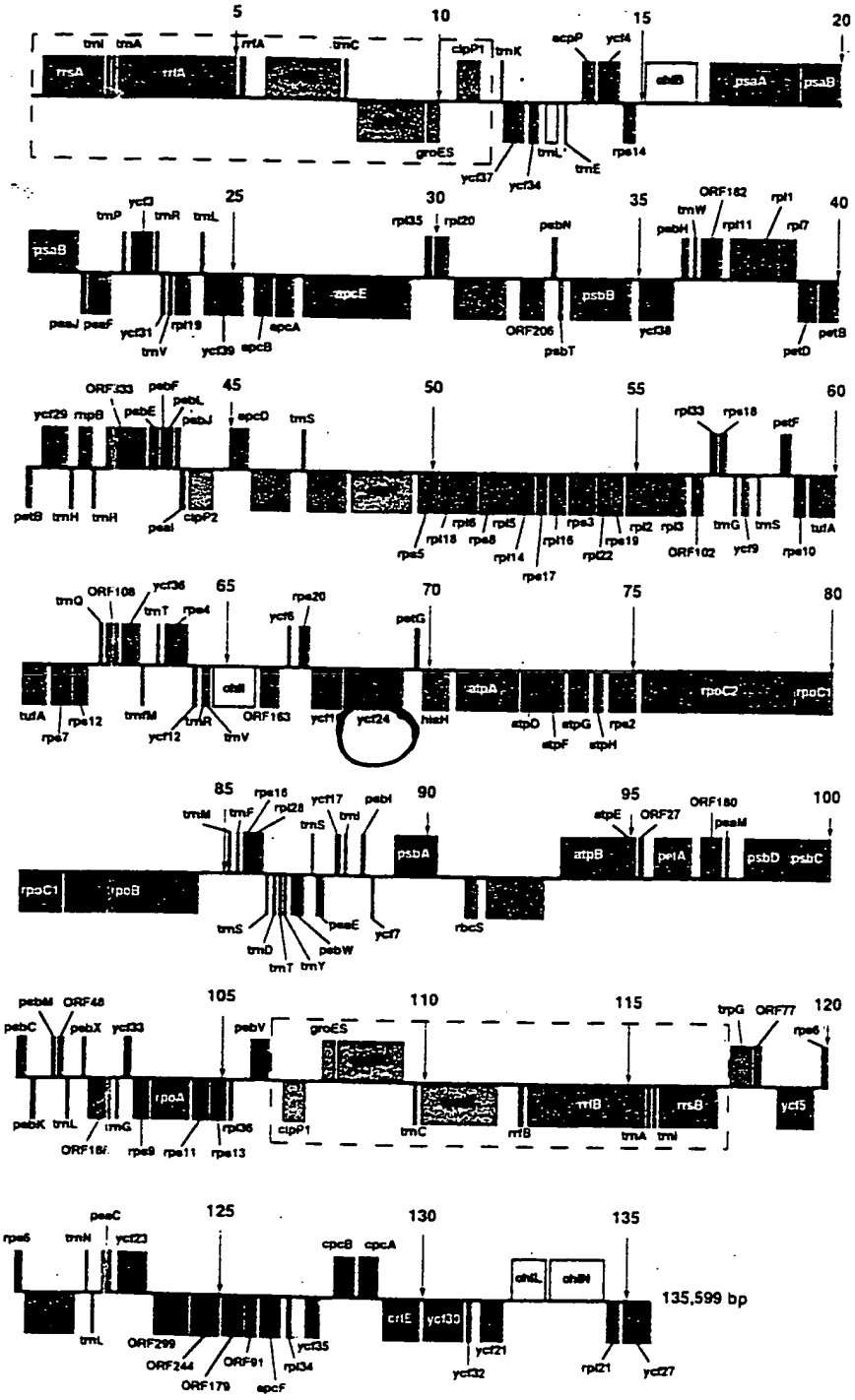
unicellular alga *Cyanophora paradoxa* Pringsheim strain LB555 has been determined (GeneBank accession number: U30821). The chromosome (Fig. 1) has a G+C content of 30.4 percent and is characterized by two inverted-repeat segments of 11,285 bp each (IR_A and IR_B, respectively), separated by a large single-copy region (LSC) of 94,946 bp and a small single-copy region (SSC) of 18,083 bp. Approximately 192 identified genes and open reading frames (of which ten are duplicated in the inverted repeats) have been identified thus far. Although the genome is smaller than that of tobacco, it encodes about 30 percent more genes than does the tobacco chloroplast genome. Genes encoded in the chloroplast genomes of higher plants but not found in the cyanelle genome include *ndhA-K*, *rps15*, *rpl23*, *infA*, *atpI*, *accD*, and *matK*.

To catalog genes and reading frames on the cyanelle DNA, the molecule has arbitrarily been linearized at the beginning of the inverted repeat 5' to the end of one of the 16S rRNA genes. Counting proceeds through IR_A, LSC, IR_B, and SSC (Fig. 1). Identified genes are given names as recommended (Hallick and Bottomley, 1983; Hallick, 1989); unknown reading frames larger than 25 codons are labeled "ORF" followed by the number of codons. Unknown reading frames that have been found in other plastid DNAs are labeled *ycf* (Hallick and Bairoch, 1993). Genes for tRNAs are identified by the amino acid with which they are charged in the single-letter code.

While a full discussion of the sequence characteristics will be presented elsewhere, we should note the special features of cyanelle DNA:

- a single type-I intron located in *trnL*(UUA) in a position that is conserved in many cyanobacteria and all plastid DNAs (Kuhsel et al., 1990; Xu et al., 1990);
- a full complement of tRNA genes (36 in total) and the gene *rnpB* (= RNA subunit of RNAseP);
- one set of rRNA genes (16 S, 23 S, and 5 S, with *trnA* and *trnI* genes in the 16 S-23 S spacers) in each IR;
- 37 ribosomal protein genes (18 small-subunit proteins; 19 large-subunit proteins);
- four genes encoding RNA polymerase subunit proteins, and *tufA* encoding translation factor Tu;
- seven genes encoding phycobiliproteins (including all chromophorylated subunits of the phycobilisome);

Fig. 1: Physical map and gene map of the cyanelle genome of *Cyanophora paradoxa* (opposite page). See table of gene symbols on p. 324.


5
10
15
20
trnI trnA
rrfA
trnC
clpP1 trnK
ycf4
rrnA
rrnA
psaA psaB
groES
ycf37
ycf34
trnL
trnE
rps14
25
30
35
40
ycf3
trnP trnR trnL
rpl35 rpl20
psbN
ORF182
trnW
psbH
rpl11
rpl1
rpl7
psaB
psaJ psaF
ycf31
trnV
rpl19
ycf39
apcB
apcA
apcE
ORF206
psbT
psbB
ycf38
petD
petB
45
50
55
60
ORF133
pebF
pebL
pebJ
ycf29 rnpB pebE
apcD
trnS
rpl33 rps18
petF
petB trnH trnH
psaI
clpP2
rps5
rpl18 rps8
rpl6 rpl5
rpl14
rpl16
rps17
rps3 rpl2 rpl3
rpl22 rps19
ORF102
trnG trnS
ycf9
rps10
tufA
65
70
75
80
trnQ
ORF108
ycf36
trnT rps4
rps20
ycf6
petG
atpA
rpoC2
rpoC1
tufA
rps12
rps7
trnfM
ycf12
trnR
trnV
ORF163
ycf1
ycf24
hisH
atpO
atpG
rps2
atpF
atpH
85
90
95
100
trnM trnF
rps16
rpl28
ycf17
pebI
trnS
trnI
psbA
atpE
ORF27
ORF180
psaM
atpB
petA
psbD psbC
rpoC1
rpoB
trnS
trnD
trnT trnY
psaE
ycf7
rbcS
105
110
115
120
pebM ORF48
pebC
pebX ycf33
pebV
groES
trpG ORF77
rps6
pebK trnL
ORF184
trnG
rps9 rps11 rps13
rpoA
rpl36
clpP1
trnC
rrfB
rrfB
rrsB
trnA trnI
ycf5
125
130
135
psaC
rps6
trnN
ycf23
cpcB cpcA
ohtL
ohtN
135,599 bp
trnL
ORF299
ORF244
ORF178
ORF91
apcF
rpl34
ycf35
crtE ycf30
ycf21
ycf32
rpl21
ycf27

- 25 genes encoding subunit proteins of the photosystem I complex (8 proteins) and photosystem II complex (17 subunits);
- seven genes encoding ATPase subunit proteins;
- six genes encoding subunits of the cytochrome b_6f complex;
- the *petF* gene encoding type I [2Fe-2S] ferredoxin;
- genes for the molecular chaperones GroEL, GroES, and DnaK located within IR segments; in addition gene *secY* is present, encoding a subunit of the preprotein translocation machinery;
- two different genes (*clpP1* and *clpP2*) encoding ClpP subunits of the Clp protease;
- one gene with homology to *ftsW*, possibly involved in cyanelle division or cell wall biosynthesis;
- 13 identified genes encoding proteins for functions in metabolism;
- cyanobacterial L_8S_8 RuBisCO encoded by an *rbcL-rbcS* operon;
- three ORFs with possible functions in the transcriptional regulation of gene expression (two OmpR family members; one LysR family member);
- 15 ORF and 21 *ycf* (ranging from 27 to 333 codons);
- five ORF/*ycf* with possible functions as membrane transporters;
- an ORF (*ycf17*) encoding a protein of 49 amino acids with strong homology to all members of the CAB/ELIP/HLIP protein superfamily;
- no *ndh* genes or pseudo-*ndh* reading frames (genes for NADH dehydrogenase).

The sequences of four algal plastid DNAs are or will soon be available (Table I): those of the cyanelle genome of *Cyanophora paradoxa* described here, the 191,028-bp plastid DNA from the red alga *Porphyra purpurea* (Reith, 1995; Reith & Munholland, 1995), the 119,704-bp plastid genome of the diatom *Odontella sinensis* (class *Bacillariophyceae*) (Kowallik et al., 1995), and the 143,170-bp plastid genome of *Euglena gracilis* (Hallick et al., 1993). These sequences complement those of chloroplast DNAs from six higher plant species: the liverwort *Marchantia polymorpha* (Ohyama et al., 1986); the gymnosperm *Pinus thunbergii* (Wakasugi et al., 1994); the monocots *Oryza sativa* (Hiratsuka et al., 1989) and *Zea mays* (140,386 bp; Maier et al., 1995); and the dicots *Nicotiana tabacum* (Shinozaki et al., 1986) and *Epifagus virginiana*, a parasitic, nonphotosynthetic species (Wolfe et al., 1992).

Acknowledgments. We wish to thank Hans Kössel, Klaus Kowallik, and Michael Reith for sharing unpublished information. This work has primarily been supported by a grant from the U.S. Department of Agriculture (National Research Initiative—Plant Genome Program) to H.J.B. and D.A.B. Additional support was

Table I. Fully sequenced plastid genomes.

Plant Species	Genome Size (bp)	Reference
Algae		
Cyanophora paradoxa	135,599	this article
Porphyra purpurea	119,704	Reith, 1995; Reith & Munhollland, 1995
Odontella sinensis	119,680	Kowallik et al., 1995
Euglena gracilis	143,170	Hallick et al., 1993
Chlorella ellipsoidea	155,000	M. Sugiura, personal communication
Higher Plants		
Marchantia polymorpha	121,024	Ohyama et al., 1986
Pinus thunbergii	119,707	Wakasugi et al., 1994
Nicotiana tabacum	155,844	Shinozaki et al., 1986
Epifagus virginiana	70,028	Wolfe et al., 1992
Oryza sativa	134,525	Hiratsuka et al., 1989
Zea mays	140,386	Maier et al., 1995

provided by the National Science Foundation (DMB-8818997; D.A.B.), the Arizona Agricultural Experiment Station (H.J.B.), and the Austrian Research Council (W.L.).

References

Hallick, R.B. 1989. Proposals for the naming of chloroplast genes. II. Update to the nomenclature of genes for thylakoid membrane polypeptides. Plant Mol. Biol. Reptr. 7:266-275.

Hallick, R.B. and A. Bairoch. 1994. Proposals for the naming of chloroplast genes. III. Nomenclature for open reading frames encoded in chloroplast genomes. Plant Mol. Biol. Reptr. 12:S29-30.

Hallick, R.B. and W. Bottomley. 1983. Proposals for the naming of chloroplast genes. Plant Mol. Biol. Reptr. 1:38-43.

Hallick, R.B., L. Hong, R.G. Drager, M.R. Favreau, A. Montfort, B. Orsat, A. Spielmann, and E. Stnitz. 1993. Complete sequence of the *Euglena gracilis* chloroplast DNA. Nucleic Acids Res. 21:3537-3544.

Hiratsuka, J., H. Shimada, R. Whittier, T. Ishibashi, M. Sakamoto, M. Mori, C. Kondo, Y. Honji, C. Sun, B. Meng, Y. Li, A. Kanno, Y. Nishzawa, A. Hirai, K. Shinozaki and M. Sugiura. 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes account for a major plastid inversion during the evolution of cereals. Mol. Gen. Genet. 217:185-194.

Kowallik, K.V., B. Stoebe, I. Schaffran, and U. Freier. 1995.The chloroplast genome of a chlorophyll *a+c*-containing alga, *Odontella sinensis*. Plant Mol. Biol. Reptr. 13:226-342

Kuhsel, M.G., R. Strickland and J.D. Palmer. 1990. An ancient group I intron shared by eubacteria and chloroplasts. Science 250:1570-1573.

Löffelhardt, W. and H.J. Bohnert. 1994. Molecular biology of cyanelles. In: *The Molecular Biology of Cyanobacteria*, (ed., D.A. Bryant), pp. 65-89, Kluwer Academic Publ., Dordrecht.

Maier, R.M., K. Neckermann, G.I. Igloi and H. Kössel. 1995. Complete sequence of the maize chloroplast genome: Gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. J. Mol. Biol. 251:614-628.

Ohyama, K., H. Fukuzawa, T. Kohchi, H. Shirai, T. Sano, S. Sano, K. Umesono, Y. Shiki, M. Takeuchi, Z. Chang, S.-i. Aota, H. Inokuchi, and H. Ozeki. 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha*. Nature 322:572-574.

Reith, M. 1995. Molecular biology of rhodophyte and chromophyte plastids. Annu. Rev. Plant Physiol. Plant Mol. Biol. 46:549-575.

Reith, M. and J. Munholland. 1995. Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. Plant Mol. Biol. Reptr. 13:333-345.

Shinozaki, K., M. Ohme, M. Tanaka, T. Wakasugi, N. Hayashida, T. Matsubayashi, N. Zaita, J. Chunwongse, J. Obokata, K. Yamaguchi-Shinozaki, C. Ohto, K. Torazawa, B.Y. Meng, M. Sugita, H. Deno, T. Kamogashira, K. Yamada, J. Kusuda, F. Takaiwa, A. Kato, N. Tohdoh, H. Shimada and M. Sugiura. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: Its gene organization and expression. EMBO J. 5:2043-2049.

Wakasugi, T., J. Tsudsuki, S. Ito, K. Nakashima, T. Tsudsuki and M. Sugiura. 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798.

Wolfe, K.H., C.W. Morden and J.D. Palmer. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc. Natl. Acad. Sci. U.S.A. 89:10648-10652.

Xu, M.Q., S.D. Kathe, H. Goodrich-Blair, S.A. Nierzwicki-Bauer and D.A. Shub. 1990. Bacterial origin of a chloroplast intron: Conserved self-splicing group I introns in cyanobacteria. Science 250:1566-1570.

Protist, Vol. 149, 51–59, February 1998 © Gustav Fischer Verlag

Protist

ORIGINAL PAPER

Evidence for a Single Origin of the 35 kb Plastid DNA in Apicomplexans

Paul Denny[a], Peter Preiser[b], Don Williamson[b], and Iain Wilson[b,1]

[a]Department of Biochemistry, Imperial College of Sience, Technology and Medicine, London, SW7 2AZ
[b]National Institute for Medical Research Mill Hill, London NW71AA

Submitted October 24, 1997; Accepted January 5, 1998
Monitoring Editor: Micheal Melkonian

Gene organization on three selected parts of the 35-kb plastid DNA of the malaria parasite *Plasmodium falciparum* was compared with that of two other apicomplexans, namely *Toxoplasma gondii* and *Eimeria tenella*. This comparison included the characteristic inverted ribosomal RNA repeat. A short segment of DNA from *Theileria annulata* also was included in a separate comparison. Criteria such as the presence or absence of particular genes, their map positions and their sequences, were used to assess whether the apicomplexan plastid DNAs originated from a single origin (a unitary hypothesis for the entire phylum), or whether disparate multiple events were more likely. The results provisionally favour a single origin although clearly this comparison of the apicomplexan plDNAs is still fragmentary. Contrary to the tendency towards homogeneity, evidence was found that the coccidian plastids may have evolved a suppressor mechanism for UGA stop codons.

Introduction

Traditional systematic observations, as well as phylogenetic studies of small subunit rRNA genes generally agree that apicomplexans are monophyletic (Levine 1987; Van de Peer et al. 1996). Distinct clades correspond to plasmodia, piroplasms, sarcocystans, coccidia etc. (Barta et al. 1991; Gajadhar et al. 1991; Escalante and Ayala 1995). In the present study we have tried to determine by comparative sequence analysis and gene organization whether apicomplexan plastid (pl) DNA originates from a single source.

An estimate of evolutionary rates for rDNAs places the divergence of apicomplexans from dinoflagellates ~900 million years (My) ago in the precambrian period, before the evolutionary radiation of higher plants and animals (Escalante and Ayala 1995). From our sequence of the 35 kb plDNA of *Plasmodium falciparum* (Wilson et al. 1996) we conjectured it to be of algal origin, acquired through secondary endosymbiosis by an ancient free-living progenitor, possibly a dinoflagellate or some related alveolate (Williamson et al. 1994). Following the principle of Occam's razor, we postulated that when the free-living progenitor became parasitic the algal plastid adopted its present form, the physiological specialization accompanying this change of life style resulting in massive reduction of the algal plastid genome to its vestigial, non-photosynthetic form (Wilson et al. 1994). According to this proposal, all extant apicomplexan lineages should carry plDNA with substantially the same gene content and linkages. An alternative scenario is that the ancestral plastid genome degenerated over a prolonged period of parasitism, allowing lineage-specific differ-

[1]Corresponding author;
fax 44-181 906 4477;
e-mail rwilson@nimr.mrc.ac.uk

ences in gene content or gene organization to emerge with apicomplexan speciation. A third possibility is that each lineage originated independently from separate photosynthetic ancestors. Although this last idea would most likely predict substantial differences in the plastid DNAs of each apicomplexan lineage, this outcome might be countered by convergent evolution, as with mitochondrial DNAs which, although widely different in organization and possibly of multiple origins, tend to have a similar genetic content (Gray and Spencer 1996).

One piece of information already tends to support the first of the three hypotheses outlined above; namely, that the plDNA of *Toxoplasma gondii* resembles that of *Plasmodium*, both in its size and propensity to form a large distinctive cruciform structure (Borst et al. 1984; Williamson et al. 1985) due to the presence of a rDNA inverted repeat with a novel gene arrangement (Wilson et al. 1993; Kohler et el. 1997). In addition, D. Roos and colleagues (personal communication) have found other genes similar to those on the malarial plDNA (the largely unannotated *T. gondii* sequence has the accession number U87145).

To assess the three hypotheses on the origin of the plDNA further, we have generated our own plDNA sequences from two coccidians, *T. gondii* and *Eimeria tenella*, as well as the piroplasm *Theileria annulata* and compared them with those of *Plasmodium falciparum*.

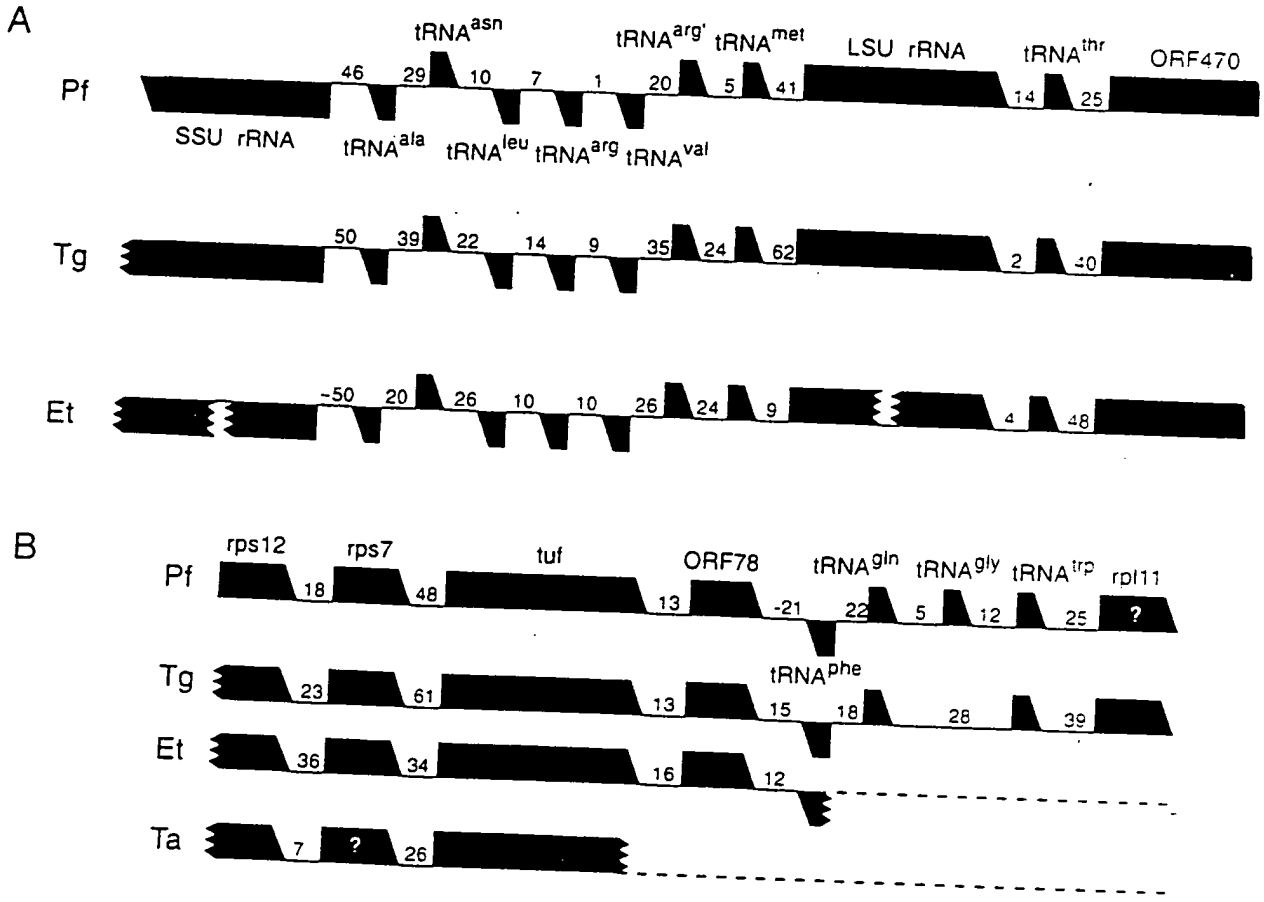


Figure 1A. Diagram (not to scale) showing conservation of gene order on the plastid-like DNAs of *Plasmodium falciparum* (Pf), *Toxoplasma gondii* (Tg), and *Eimeria tenella* (Et). In Pf this region of the plDNA spans approximately 6.3 kb. It includes one copy of the rDNA region of the inverted repeat and genes immediately downstream of the large subunit rRNA (LSU rRNA) – see Wilson et al. (1996). The lengths of intergenic regions are given as the number of nucleotides. Arginyl tRNA genes with different anticodons are distinguished.

Figure 1B. Diagram (not to scale) showing the gene arrangement on either side of the *tuf* gene on the plDNAs of *P. falciparum* (Wilson et al. 1996), *T. gondii*, *E. tenella* and *Th. annulata* (Ta). In Pf this region spans ~3.0 kb of the plDNA. As indicated, the $tRNA^{phe}$ gene is on the opposite strand from *tuf* in all cases examined. The primers used for PCR did not extend to the dashed regions.

Results

Gene Organization

Three regions of the *Plasmodium* 35 kb plDNA with characteristic gene arrangements were selected, and PCR products covering these regions were amplified from three other apicomplexan DNAs. The amplicons were cloned, sequenced and compared with *Plasmodium* plDNA for genetic content, gene organization, and in some cases for nucleotide sequence similarity.

The Inverted Repeat

The inverted repeat (IR) in *Plasmodium* comprises a novel arrangement of duplicated small and large rRNA genes interspersed with nine tRNA genes, occupying both strands on this sector of the plDNA (Gardner et al. 1994). Earlier "snap-back" back experiments with DNA from *Toxoplasma* and *Eimeria* were consistent with the presence of an IR and sequence data obtained in the present study confirmed the finding of Roos and colleagues (Accession No. U87145) that the same gene order is conserved (Fig. 1A). It may be noted in passing that a small subunit rRNA gene with similar sequence has also been reported from *Babesia bovis* (Gozar and Bagnara, 1995).

The ORF470 Region

ORF470 corresponds to the highly conserved gene *ycf 24* that has a limited distribution in known plastids. Our findings, summarized in Fig. 1A, show that in both *Toxoplasma* and *Eimeria* a tRNA$^{thr(ugu)}$ gene

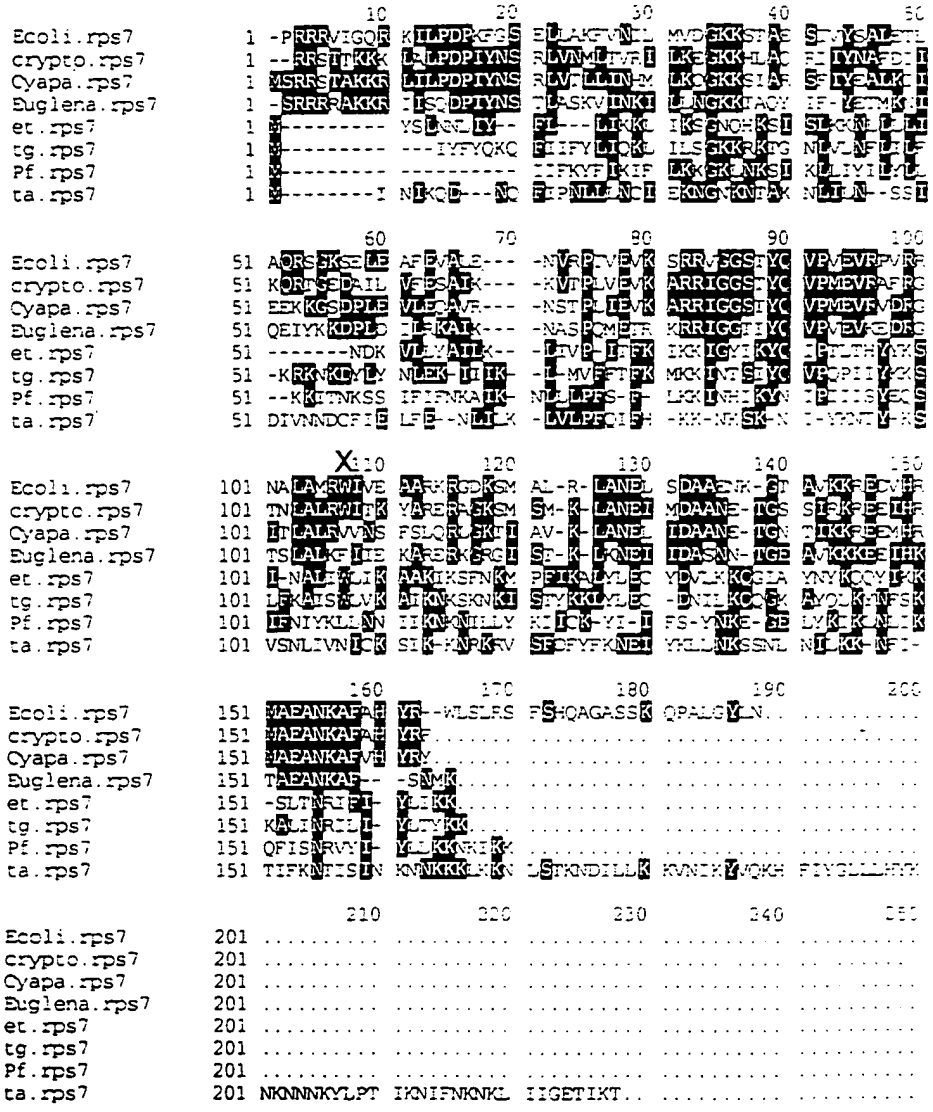


Figure 2. An alignment of the putative rps 7 sequences from four apicomplexans (Pf, Tg, Et, and Ta) and those from *Escherichia coli* (V00355) and the plDNAs of *Cryptomonas phi* (X52912), *Cyanophora paradoxa* (X52497) and *Euglena gracilis* (X70810). A cross indicates the position of a conserved tryptophan in the *T. gondii* sequence encoded by a UGA "stop codon".

and a homologue of the *Plasmodium* ORF470 gene lie downstream of an LSU rRNA gene, just as in *Plasmodium* (Williamson et al. 1994).

The *Tuf* Region

In *Plasmodium*, the ribosomal protein genes *rps 7* and *rps* 12 precede the *tuf* gene, all three lying on the opposite strand from the adjacent, downstream gene for $tRNA^{phe}$. The same overall arrangement was found on sequenced PCR products derived from the purified plDNA of *Toxoplasma*, and total DNAs from *Eimeria* and *Theileria* (Fig. 1B). Unlike *rps* 12 which is relatively well conserved (~50% amino-acid identity with *E. coli*), the putative *rps 7* gene was poorly conserved in all the apicomplexan DNAs (Fig. 2A). Although ORF78 downstream of the *tuf* gene in *Plasmodium* also was not well conserved in *Toxoplasma* (ORF43) or *Eimeria* (ORF45), conservation of a motif in the latter allowed us tentatively to relate (32% amino acid identity) the apicomplexan sequences to ORF57 found on the plDNA of the euglenoid *Astasia longa* (Gockel et al. 1994) – see Fig. 3A. Another tentative assignment was made for ORF129 located still further downstream of *tuf* in *Plasmodium* (Wilson et al. 1996) – this time a motif in the *Toxoplasma* (ORF123) version led to its provisional identification as the ribosomal protein encoding gene *rpl* 11[BLAST score = 6.1 e-09 with the gene from *Cyanophora*] (Fig. 3B).

One minor difference in the plDNAs of *Plasmodium* and *Toxoplasma* was noted in the *tuf* region. Of the three contiguous tRNA genes in *Plasmodium* close to but on the opposite strand from $tRNA^{phe}$ (i.e. glutamine (Q), glycine (G) and trytophan (W)), only those for Q and W were present in *Toxoplasma*. Apart from this exception, the order of all the contiguous genes we provisionally identified was conserved in *Plasmodium, Toxoplasma, Eimeria* and *Theileria*.

As well as gene order, the sequences of the genes on the apicomplexan plDNAs were more similar to each other than to other plastid versions. For example, the N-terminus of the hypothetical translation product of *ycf* 24 of *Porphyra purpurea* was more similar to the cyanobacterial version (*Synechocystis* sp.) than were the apicomplexan ones (compare amino acid blocks 20–30; 30–40; 50–60; 70–80 in Fig. 4). Yet the overall homology was readily apparent over the rest of the protein.

The conserved nature of the apicomplexan protein sequences, as well as their unique gene order, support the idea that the plDNAs originated from a single source.

Tryptophan Codons

Unlike *Plasmodium*, the coccidians *Toxoplasma* and *Eimeria* appeared to use an opal stop codon (UGA) at a few sites in conserved tryptophan positions. As shown in Figure 4, this occurred at different single

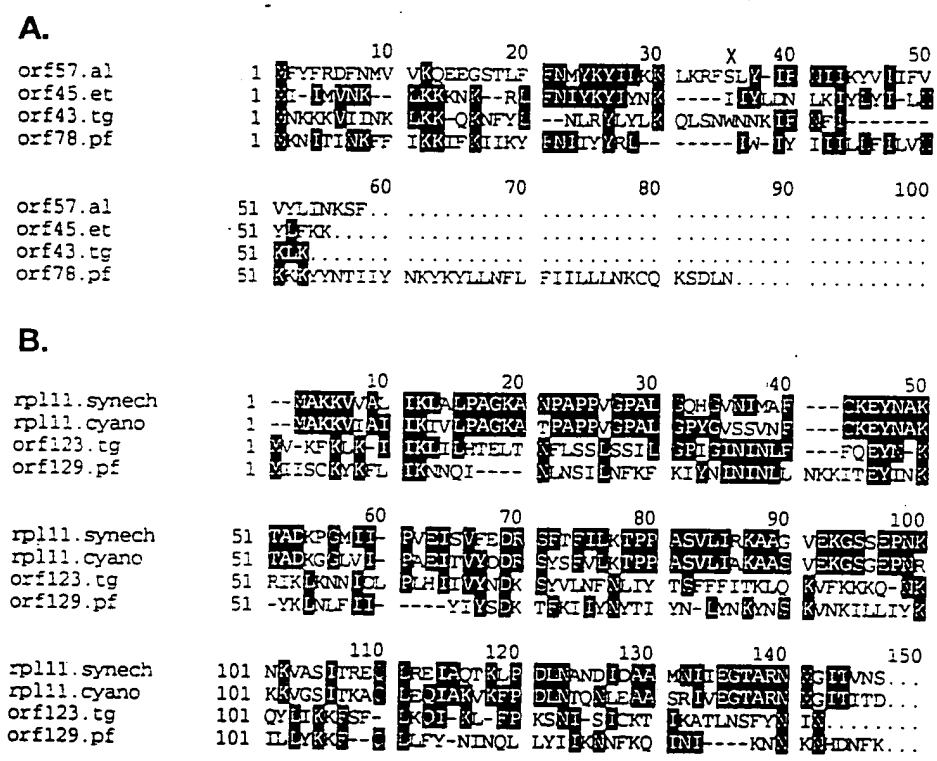


Figure 3.
A) An alignment showing similarity between ORF57 from *Astasia longa* (Gockel et al. 1994) and the small ORF lying downstream from the *tuf* gene on the apicomplexan plDNAs. A cross indicates the position of a tryptophan in the *T. gondii* sequence encoded by a UGA "stop codon".
B) An alignment of the putative apicomplexan *rpl* 11 sequences from *P. falciparum* (ORF129) and *T. gondii* (ORF123) with those of *Synechocystis* spp. (P36237) and *Cyanophora paradoxa* (P48126).

sites (cross and filled circle, respectively) in the *Toxoplasma* and *Eimeria* versions of *ycf* 24. The *Toxoplasma* opal stop codon occurred at the same site in the sequence obtained independently by Roos et al. (Accession No. U87145). In both *Toxoplasma* and *Eimeria*, the sequence around the unusual codon was highly conserved and interpretation of the opal stop as a tryptophan codon restored conserved C-terminal sequences. In the case of *rps 7* (a less well conserved protein) *Toxoplasma* again used a UGA codon in a tryptophan position that is often conserved outside the apicomplexans (Fig. 2). We found no evidence for RNA editing of any of these positions (data not shown) and infer that the coccidian plastid has evolved a mechanism for UGA stop codon suppression (see Discussion).

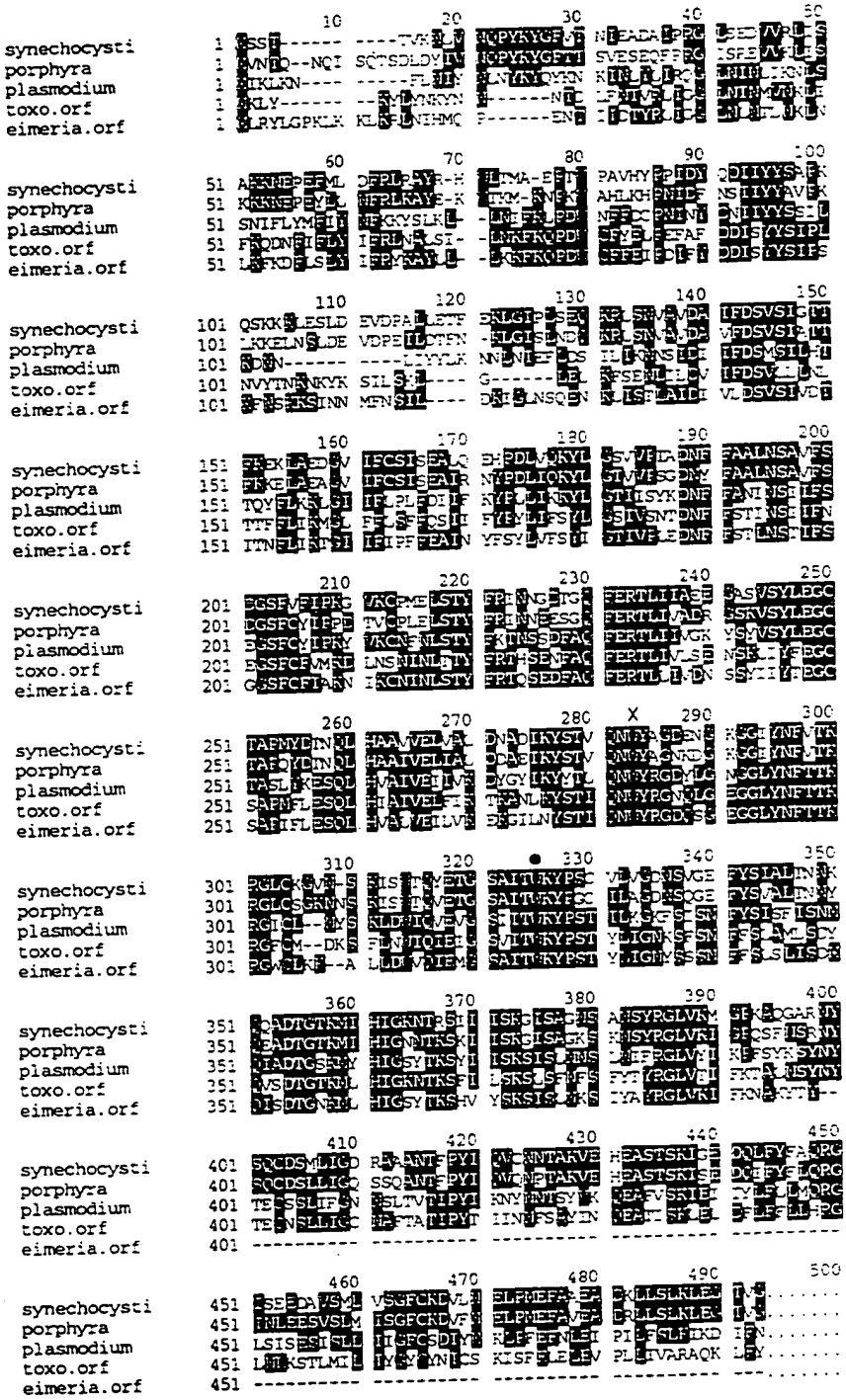


Figure 4. An alignment of the hypothetical translation products of apicomplexan versions of Pf ORF470 (*ycf* 24) with those of the plastid from *Porphyra purpurea* (U38804) and the cyanobacterium *Synechocystis* sp. (D64004). Our sequences for *T. gondii* and *E. tenella* were incomplete downstream of a conserved region used to design the PCR primers at the C-terminus. For the purposes of this figure, the region downstream of position 398 was filled-in for *T. gondii* using the database entry (U87145) of Roos and colleagues. Tryptophans predicted from UGA "stop codons" are indicated by a cross for *T. gondii* and filled circle for *E. tenella*.

Intergenic Regions

Intergenic regions of the apicomplexan plDNA were extremely A·T-rich and more heterogeneous than the open reading frames, indels of different length being found. But because intergenic regions (other than control sequences) are less likely to be constrained by natural selection than functional open reading frames, these findings are probably not relevant to the hypothesis being examined, namely whether all the plDNAs evolved from a single source.

Discussion

We have shown that portions of the plDNA of four apicomplexan genera (covering about one third of the total length of *Plasmodium* plDNA), have a highly conserved gene content, gene order and gene sequences. This favours the hypothesis that they evolved from a single photosynthetic progenitor. The 35 kb plDNA is about one fifth the size of a conventional plastid DNA molecule and to have reached its present form and novel gene arrangement it must have undergone numerous deletions and rearrangements. As the apicomplexans are believed to have diverged as long ago as 800 My (Escalante and Ayala 1995), it seems unlikely their plDNAs became reduced to the present novel arrangement through convergent evolution from different origins. It may be noted in this regard that the vestigal plDNAs of two parasitic higher plants, *Epifagus virginiana* (beechdrops) (Wolfe et el. 1992a) and *Conopholis americana* (squawroot) (Wimpee et el. 1992) from the family Orobanchaceae (of the parasitic branch of the Scrophulariales), now differ considerably in gene content and organization as the result of extensive deletions and rearrangements incurred since they diverged from their congeners over the much shorter period of some 5–50 My (Muller, 1981; Wolfe et al. 1992b).

Like ourselves, D. Roos and his colleagues have found UGA "stop codons" at a few conserved tryptophan residues in the plDNA of *Toxoplasma* (personal communication). Thus it would be doubly interesting to confirm translation of the gene products of *ycf* 24 and *rps* 7 and to clarify the proposed suppressor mechanism for these UGA stop codons. Unfortunately, at present the reagents required to demonstrate protein expression (for example, specific antibodies) are unavailable. A number of different suppressor mechanisms for stop codons have been recorded in other systems; these include mutations in the peptidyl-transferase region of the large subunit rRNA (Jemiolo et el. 1995), and specific deletions in the small subunit rRNA (Murgola et el. 1988). Neither of these seems applicable in the present case (Denny, unpublished data). Likewise, various mutations in tryptophan tRNAs can act as suppressors of UGA stop codons (Hirsh 1971; Hirsh and Gold 1971). This list can be extended to release factors and ribosomal proteins, but for the present we have no clear indication of how suppression might be achieved in the coccidial plastid organelle and further work will be required to clarify this issue.

A recent phylogenetic analysis of the apicomplexan plastid *tuf* gene (Kohler et el. 1997) placed it with green plastids rather than with those of the red algae we originally suggested might be closest based on the then known distribution of ORF470 (*ycf* 24) (Williamson et al. 1994). But bootstrap support among the plastid clades is weak in such analyses and so there is as yet no clear support for any one plastid group furnishing the potential endosymbiont whose plDNA has been maintained in a vestigial form in apicomplexans.

The highly conserved nature of the gene content of the vestigial plDNA in four genera of apicomplexans, despite their evolution of different life styles over a long evolutionary time span, suggests the organelle has an essential function (McFadden and Waller 1997; Kohler et al. 1997). Direct evidence for apicomplexan plastid proteins remains to be gathered, but transcriptional studies have confirmed that the *rRNA*, *tRNA*, *tuf*, *ycf* 24, and *rps* 7 genes mentioned here are all transcribed in erythrocytic forms of *Plasmodium* and tachyzoites of *Toxoplasma* (Wilson et al. 1996, and data not shown).

To strengthen our postulate that the plDNA of apicomplexans comes from a single source, information should be gathered from other members of the phylum, especially those that could be considered less specialized, such as the gregarines which infest invertebrate lineages. For the present, however, we conclude that apicomplexan parasites of vertebrates, despite their extensive evolutionary divergence, originated from a single photosynthetic ancestral cell whose plDNA underwent massive deletion and rearrangement, either leading to or following the parasitic event.

Methods

DNA. All the DNA sequences reported in this paper were obtained from the following starting materials. Total DNA from *P. falciparum* – clone C10 (Hempelmann et al. 1981) was extracted and fractionated on caesium chloride/4', 6-diamidino-2-phenylindole (DAPI) gradients to yield isolated 35 kb DNA, as pre-

Table 1. Oligonucleotide primers for amplification of apicomplexan plastid DNAs.

Segment of circle	PCR primers pairs
a. Toxoplasma gondii	
rRNA/tRNA Inverted Repeat-ORF470	5' TACGGCTACCTTGTTACGACTTC and 5' GCGGTAATACAGAAAATGCAAG[1]
	5' AAGAACATACCAATCCACC and 5' CTATTTCACCGAATATTTC
	5' GCGAAATTCCTTGTCGGGTAAGTTCC[2] and 5' TTTT/CT/CGG/ATCCTCTCGTAC[2]
	5' GTTCGCCTATTAAAGCGATACGTGAGCTGGG and 5' ATCACATTGAG/CA/TATAATT
rps 12-rps 7-tuf-tRNAs-rpl 11	5' GGAGGTAGAGTAAAAGATTTACCAGG and 5' GGTAGAGCAATGGATTGAAG
	5' CTTCAATCCATTGCTCTACC and 5' GTGAGTATAGTTTAG
b. Eimeria tenella	
rRNA/tRNA Inverted Repeat-ORF470	5' CTGAGCCAGGATCAAACTC and 5'GGAAAACTGCTTCTAAG
	5' CTTAGAAGCAGTTTTCC and 5' TTTCTCACTTATATGTTGTC
	5' GTTCGCCTATTAAAGCGATACGTGAGCTGGG and 5' ATCACATTGAG/CA/TATAATT
rps 12-rps 7-tuf-tRNA^{Pro}	5' GGAGGTAGAGTAAAAGATTTACCAGG and 5' GGTAGAGCAATGGATTGAAG
c. Theileria annulata	
rps 12-URF-tuf	5' GGAGGTAGAGTAAAAGATTTACCAGG and 5' ATCTTCTTTATTTAAAAAA/TAC

[1] Gozar and Bagnara (1995).
[2] Beckers et al. (1995).

viously described (Gardner et al. 1988). *T. gondii* (*Toxoplasma* – RH strain) was obtained either from infected cotton rats (courtesy of Dr. Judith Smith, Leeds University) or from infected cultures of the dog epithelial cell line MDCK. Following buoyant density centrifugation of total *Toxoplasma* DNA as described above, the 35 kb DNA homologue was identified as a low density satellite band. This was removed by side puncture of the centrifuge tube, purified as a pellet by further centrifugaton, and shown to cross-hybridize in Southern blots with probes from the *Plasmodium* 35 kb DNA. Total DNA was also extracted from sporozoites of *Eimeria tenella* – the H strain was supplied by Dr Fiona Tomley, Institue of Animal Health, Compton, and from trophozoites of *Theileria annulata* in bovine blood – the Hissar strain was supplied by Dr Roger Hall, Dept. Biology, York University.

Polymerase chain reaction (PCR). Oligonucleotide primers were prepared based on conserved regions of the *Plasmodium* 35 kb plDNA or from other published sequences (Table 1). PCR products were generated using either the *Toxoplasma* circular DNA or, in the cases of *Eimeria* and *Theileria* total genomic DNAs as template. PCR reactions were carried out with 20 ng *Toxoplasma* plDNA, 150 ng total *Eimeria* DNA, or 300 ng total *Theileria* DNA as templates. Approximately 50 pmol of each primer was added to reaction mixtures containing 2.5 mM $MgCl_2$, x1 PCR buffer (Promega) and 1 µl (5 units) Amplitaq (Cetus) in 100 µl total volume. Amplification was carried out for 35 cycles as follows: 95 °C for 30 sec, 40 °C for 60 sec, 72 °C for 120 sec. As negative controls, total DNA was used from *Leishmania guyanensis* B8 (courtesy Dr Douglas Barker, Molteno Laboratories, Cambridge), or from MDCK cells.

Southern Hybridization. DNAs were digested at 37 °C overnight with *Hin*dIII (Gibco/BRL), and the restriction fragments from approximately 1 µl of total DNA from *P. falciparum*, *T. gondii*, *E. tenella* and *Th. annulata*, as well as 20 ng of isolated *P. falciparum* plDNA, were separated by electrophoresis in 0.8% agarose in the presence of 0.6 µg ml^{-1} Ethidium Bromide. The DNA was transferred under alkaline conditions to Hybond N+ membranes, according to the manufacturer's instructions and hybridization was carried out using ^{32}P-labelled probes prepared by the random primer method (as described in the Prime-It II Kit, Stratagene). In some experiments a lower stringency was used for hybridization and washing (45–50 °C with high salt).

Sequence analysis. PCR products were purified using the Wizard PCR Preps kit (Promega) and sequenced on one strand by the Sanger di-deoxy chain termination method (Sambrook et al. 1989) using the Sequenase Version 2.0 kit (USB). If necessary, terminal dideoxy-transferase (TdT) was included, as described previously (Fawcett and Bartlett, 1990). The sequences were assembled using the Staden-Plus programme SAP (Amersham). Reverse strands were sequenced by fluorescent dye-terminated thermo-cycle sequencing (Perkin Elmer) using an ABI PRISM 377 automated DNA sequencer (Perkin Elmer). Sequences were further analysed using the Sequence Analysis Software Package (Genetics Computer Group, Wisconsin version 7.0) (Devereux et al. 1984). EMBL database accession numbers for data presented here are as follows: *Plasmodium* X95275, X95276; *Toxoplasma* Y11430, Y11431; *Eimeria* Y12332, Y12333; *Theileria* Y11429.

Acknowledgements

We are grateful to Dr Judith Smith, Leeds University; Dr Mary Pudney, Glaxo/Wellcome; Dr Fiona Tomley, Institue of Animal Health, Compton and Dr R. Hall, Dept. Biology, York University, U.K. for providing *Toxoplasma* isolates and *Eimeria* and *Theileria* materials. We also thank Irene Ling and Malcolm Strath (N.I.M.R.) for carrying out the automated sequencing work. P.D. was the recipient of an British MRC Studentship award. This work was also supported by the UNDP/World Bank/WHO Special Progamme for Research in Tropical Diseases (TDR).

References

Barta JR, Jenkins MC, Danforth HD (1991) Evolutionary relationships of avian *Eimeria* species among other Apicomplexan protozoa: Monophyly of the Apicomplexa is supported. Mol. Biol Evol **8**: 345–355

Beckers CJM, Roos DS, Donald RGK, Luft BJ, Schwab JC, Yang C, Joiner KA (1995) Inhibition of cytoplasmic and organellar protein synthesis in Toxoplasma gondii: Implications for the target of macrolide antibiotics. J Clin Invest **95**: 367–376

Borst P, Overdulve JP, Weijers PJ, Fase-Fowler F, van den Berg M (1984) DNA circles with cruciforms from *Isospora (Toxoplasma) gondii*. Biochem Biophys Acta **781**: 100–111

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programmes for the VAX. Nucl Acids Res **12**: 387–395

Escalante AA, Ayala FJ (1995) Evolutionary origin of *Plasmodium* and other Apicomplexa based on rRNA genes. Proc Natl Acad Sci USA **92**: 5793–5797

Fawcett TW, Bartlett SG (1990) An effective method for eliminating "artifact banding" when sequencing double-stranded DNA template. Biotechniques **9**: 46–48

Gajadhar AA, Marquardt WC, Hall R, Gunderson JG, Aritzia-Carmona EV, Sogin ML (1991) Ribosomal RNA sequences of *Sarcocystis muris, Theileria annulata* and *Crypthecodinium cohnii* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. Mol Biochem Parasitol **45**: 147–154

Gardner MJ, Bates PA, Ling IT, Moore DJ, McCready S, Gunasekera MBR, Wilson RJM, Williamson DH (1988) Mitochondrial DNA of the human malarial parasite *Plasmodium falciparum*. Mol Biochem Parasitol **31**: 11–18

Gardner MJ, Preiser P, Rangachari K, Moore D, Feagin JE, Williamson DH, Wilson RJM (1994) Nine duplicated tRNA genes on the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. Gene **140**: 307–308

Gockel GW, Hachtel W, Baier S, Fliss K, Henke M (1994) Genes for components of the chloroplast translational apparatus are conserved in the reduced 73-kb plastid DNA of the nonphotosynthetic euglenoid flagellate *Astasia longa*. Curr Genet **26**: 256–262

Gozar MMG, Bagnara AS (1995) An organelle-like small subunit ribosomal RNA gene from *Babesia bovis*: Nucleotide sequence, secondary structure of the transcript and preliminary phylogenetic analysis. Int J Parasitol **25**: 929–938

Gray MW, Spencer DF (1996) Organellar evolution. In Roberts DMcL, Sharp PM, Alderson C, Collins M (eds) Evolution of Microbial Life, Society for General Microbiology **54**, University of Warwick, Cambridge University Press pp. 109–126

Hempelmann E, Ling I, Wilson RJM (1981) S-antigens and isozymes in strains of *Plasmodium falciparum*. Trans R Soc Trop Med Hyg **75**: 855–858

Hirsh D (1971) Tryptophan transfer RNA as the UGA suppressor. J Mol Biol **58**: 439–458

Hirsh D, Gold L (1971) The translation of the UGA triplet in vitro by tryptophan transferase RNAs. J Mol Biol **58**: 459–468

Jemiolo DK, Pagel FT, Murgola EJ (1995) UGA suppression by a mutant RNA of the large ribosomal subunit. Proc Natl Acad Sci USA **92**: 12309–12313

Kohler S, Delwiche CF, Denny P, Tilney LG, Webster P, Wilson RJM, Palmer JD, Roos D (1997) A plastid of probable green algal origin in apicomplexan parasites. Science **275**: 1485–1489

Levine ND Ed. (1991) Phylum II. Apicomplexa. An Illustrated Guide to the Protozoa. Kansas, Lawrence: Society of Protozoologists pp. 322–357

McFadden GI, Waller RF (1997) Plastids in parasites of humans. Bioessays **19**: 1033–1040

Muller J (1981) Fossil pollen records of extant angiosperms. Bot Rev **47**: 1–142

Murgola EJ, Hijazi KA, Goringer HU, Dahlberg AE (1988) Mutant 16S ribosomal RNA: A codon translational suppressor. Proc Natl Acad Sci USA **85**: 4162–4165

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular Cloning. A laboratory manual. New York, Cold Spring Harbor Laboratory Press

Van de Peer YG, Auwera VG, De Wachter R (1996) The evolution of stramenopiles and alveolates as derived by "substitution rate calibration" of small ribosomal subunit RNA. J Mol Evol **42**: 201–210

Williamson DH, Gardner MJ, Preiser P, Moore DJ, Rangachari K, Wilson RJM (1994) The evolutionary origin of the malaria parasite's 35 kb circular DNA; new evidence supports a possible rhodophyte ancestry. Mol Gen Genet **243**: 249–252

Williamson DH, Wilson RJM, Bates PA, McCready S, Perler F, Qiang B (1985) Nuclear and mitochondrial DNA of the primate malarial parasite *P. knowlesi*. Mol Biochem Parasitol **14**: 199–209

Wilson I, Gardner M, Rangachari K, Williamson D (1993) In Smith JE (ed) Extrachromosomal DNA in the Apicomplexa. Toxoplasmosis. NATO ASI series H 78 Heidelberg, Springer-Verlag pp. 51–62

Wilson RJM, Denny PW, Preiser PR, Rangachari K, Roberts K, Roy A, Whyte A, Strath M, Moore DJ, Moore PW, Williamson DH (1996) Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. J Mol Biol **261**: 155–172

Wilson RJM, Williamson DH, Preiser P (1994) Malaria and other Apicomplexans: The "plant" connection. Infect Agents Dis **3**: 29–37

Wimpee CF, Morgan R, Wrobel RL (1992) Loss of transfer RNA genes from the plastid 16S-23S ribosomal RNA gene spacer in a parasitic plant. Curr Genet **21**: 417–422

Wolfe KH, Morden CW, Palmer JD (1992a) Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc Natl Acad Sci USA **89**: 10648–10652

Wolfe KH, Morden CW, Ems SC, Palmer JD (1992b) Rapid evolution of the plastid translational apparatus in a nonphotosynthetic plant: Loss or accelerated sequence evolution of tRNA and ribosomal protein genes. J Mol Evol **35**: 304–317

APPENDIX D

Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?CMD=search&DB=nucleotide
search : ycf24

1: D64004
Synechocystis sp. PCC 6803 DNA, complete genome, section:23/27, 2868767-3002965
gi|1001701|dbj|D64004.1|SYCSLRF[1001701]

2: NC_004113
Thermosynechococcus elongatus BP-1, complete genome
gi|22297544|ref|NC_004113.1|[22297544]

3: NC_003272
Nostoc sp. PCC 7120 complete genome
gi|17227497|ref|NC_003272.1|[17227497]

4: NC_000911
Synechocystis sp. PCC 6803, complete genome
gi|16329170|ref|NC_000911.1|[16329170]

5: NC_000916
Methanobacterium thermoautotrophicum str. Delta H complete genome
gi|15678031|ref|NC_000916.1|[15678031]

6: NC_001799
Toxoplasma gondii apicoplast, complete genome
gi|11496534|ref|NC_001799.1|[11496534]

7: NC_001713
Odontella sinensis chloroplast, complete genome
gi|11467432|ref|NC_001713.1|[11467432]

8: NC_001675
Cyanophora paradoxa cyanelle, complete genome
gi|11467282|ref|NC_001675.1|[11467282]

9: NC_000925
Porphyra purpurea chloroplast, complete genome
gi|11465652|ref|NC_000925.1|[11465652]

10: NC_000926
Guillardia theta chloroplast, complete genome
gi|11467607|ref|NC_000926.1|[11467607]

11: NC_001840
Cyanidium caldarium chloroplast, complete genome
gi|11465393|ref|NC_001840.1|[11465393]

12: AP005370
Thermosynechococcus elongatus BP-1 DNA, complete genome, section 2/9
gi|22294033|dbj|AP005370.1|[22294033]

13: AE000884
Methanobacterium thermoautotrophicum from bases 1050856 to 1062059 (section 90 of 148) of the complete genome
gi|2622242|gb|AE000884.1|[2622242]

14: AP003589
Nostoc sp. PCC 7120 DNA, complete genome, section 9/19
gi|17131372|dbj|AP003589.1|[17131372]

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

15: BI437350
gc59d11.y1 Moss EST library PPN Physcomitrella patens cDNA clone
PEP_SOURCE_ID:PPN200721 5' similar to SW:YC24_ODOSI P49530 HYPOTHETICAL
54.3 KD PROTEIN YCF24 ;, MRNA sequence
gi|15262040|gb|BI437350.1|[15262040]

16: Z67753
Odontella sinensis complete chloroplast genome
gi|1185127|emb|Z67753.1|OSCHLPLXX[1185127]

17: AJ132267
Skeletonema costatum chromoplast ycf24 gene, partial
gi|4210403|emb|AJ132267.1|SCO132267[4210403]

18: AF022186
Cyanidium caldarium strain RK1 chloroplast, complete genome
gi|6466296|gb|AF022186.2|AF022186[6466296]

19: AF138960
Neospora caninum ycf24 protein (ycf24) gene, partial cds; DNA dependent RNA polymerase beta subunit (rpoB) gene, complete cds; and DNA dependent RNA polymerase beta subunit'(rpoC1) gene, partial cds, plastid genes for plastid products
gi|6492292|gb|AF138960.1|AF138960[6492292]

20: U87145
Toxoplasma gondii chloroplast, complete genome
gi|5231237|gb|U87145.2|U87145[5231237]

21: AF095904
Toxoplasma gondii ycf24 protein (ycf24) gene, partial cds; DNA dependent RNA polymerase beta subunit (rpoB) gene, complete cds; and DNA dependent RNA polymerase beta' subunit (rpoC1) gene, plastid genes encoding plastid proteins, partial cds
gi|4336507|gb|AF095904.1|AF095904[4336507]

22: AF041468
Guillardia theta complete plastid genome
gi|3602932|gb|AF041468.1|[3602932]

23: U38804
Porphyra purpurea chloroplast, complete genome
gi|1276652|gb|U38804.1|PPU38804[1276652]

24: D90812
E.coli genomic DNA, Kohara clone #321(38.1-38.4 min.)
gi|1742768|dbj|D90812.1|[1742768]

25: D90811
E.coli genomic DNA, Kohara clone #320(37.9-38.3 min.)
gi|1742754|dbj|D90811.1|[1742754]

26: U30821
Cyanophora paradoxa cyanelle, complete genome
gi|1016083|gb|U30821.1|CPU30821[1016083]

APPENDIX D

Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
1: D64004. Synechocystis sp....[gi:1001701]

LOCUS       SYCSLRF               134199 bp    DNA     linear   BCT 08-FEB-2003
DEFINITION  Synechocystis sp. PCC 6803 DNA, complete genome, section:23/27,
            2868767-3002965.
ACCESSION   D64004 AB001339 BA000022
VERSION     D64004.1  GI:1001701
KEYWORDS    .
SOURCE      Synechocystis sp. PCC 6803
  ORGANISM  Synechocystis sp. PCC 6803
            Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
REFERENCE   1
  AUTHORS   Kaneko,T., Tanaka,A., Sato,S., Kotani,H., Sazuka,T., Miyajima,N.,
            Sugiura,M. and Tabata,S.
  TITLE     Sequence analysis of the genome of the unicellular cyanobacterium
            Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
            region from map positions 64% to 92% of the genome
  JOURNAL   DNA Res. 2 (4), 153-166 (1995)
  MEDLINE   96127529
   PUBMED   8590279
REFERENCE   2
  AUTHORS   Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y.,
            Miyajima,N., Hirosawa,M., Sugiura,M., Sasamoto,S., Kimura,T.,
            Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Naruo,K.,
            Okumura,S., Shimpo,S., Takeuchi,C., Wada,T., Watanabe,A.,
            Yamada,M., Yasuda,M. and Tabata,S.
  TITLE     Sequence analysis of the genome of the unicellular cyanobacterium
            Synechocystis sp. strain PCC6803. II. Sequence determination of the
            entire genome and assignment of potential protein-coding regions
  JOURNAL   DNA Res. 3 (3), 109-136 (1996)
  MEDLINE   97061201
   PUBMED   8905231
REFERENCE   3  (bases 1 to 134199)
  AUTHORS   Tabata,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-1995) Satoshi Tabata, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:tabata@kazusa.or.jp, URL:http://www.kazusa.or.jp/cyano/,
            Tel:81-438-52-3933(ex.2330), Fax:81-438-52-3934)
COMMENT     Potential protein coding regions were assigned on the basis of
            similarity search of the ORFs and GeneMark analysis.
FEATURES             Location/Qualifiers
     source          1..134199
                     /organism="Synechocystis sp. PCC 6803"
                     /strain="PCC6803"
                     /db_xref="taxon:1148"
                     /note="synonym:Synechocystis PCC6803"
....
   gene              2908..4350
                     /gene="ycf24"
     CDS             2908..4350
                     /gene="ycf24"
                     /note="ORF_ID:slr0074"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter subunit"
                     /protein_id="BAA10542.1"
                     /db_xref="GI:1001705"
```

APPENDIX D

Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

/translation="MSSTTVKNLVNQPYKYGFVTNIEADAIPRGLSEDVVRLISAKKN
EPEFMLDFRLRAYRHWLTMAEPTWPAVHYPPIDYQDIIYYSAPKQSKKKLESLDEVDP
ALLETFEKLGIPLSEQKRLSNVAVDAIFDSVSIGTTFKEKLAEDGVIFCSISEALQEH
PDLVQKYLGSVVPTADNFFAALNSAVFSDGSFVFIPKGVKCPMELSTYFRINNGDTGQ
FERTLIIAEEGASVSYLEGCTAPMYDTNQLHAAVVELVALDNADIKYSTVQNWYAGDE
NGKGGIYNFVTKRGLCKGVNSKISWTQVETGSAITWKYPSCVLVGDNSVGEFYSIALT
NNKQQADTGTKMIHIGKNTRSIIISKGISAGNSANSYRGLVKMGPKAQGARNYSQCDS
MLIGDRAAANTFPYIQVDNNTAKVEHEASTSKIGEDQLFYFAQRGISEEDAVSMLVSG
FCKDVLNELPMEFAAEADKLLSLKLEGTVG"

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
2: NC_004113. Thermosynechococc...[gi:22297544]

Links


LOCUS       NC_004113            2593857 bp    DNA     circular BCT 10-DEC-2002
DEFINITION  Thermosynechococcus elongatus BP-1, complete genome.
ACCESSION   NC_004113
VERSION     NC_004113.1  GI:22297544
KEYWORDS    .
SOURCE      Thermosynechococcus elongatus BP-1
  ORGANISM  Thermosynechococcus elongatus BP-1
            Bacteria; Cyanobacteria; Chroococcales; Thermosynechococcus.
REFERENCE   1  (bases 1 to 2593857)
  AUTHORS   Nakamura,Y., Kaneko,T., Sato,S., Ikeuchi,M., Katoh,H., Sasamoto,S.,
            Watanabe,A., Iriguchi,M., Kawashima,K., Kimura,T., Kishida,Y.,
            Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A., Nakazaki,N.,
            Shimpo,S., Sugimoto,M., Takeuchi,C., Yamada,M. and Tabata,S.
  TITLE     Complete genome structure of the thermophilic cyanobacterium
            Thermosynechococcus elongatus BP-1
  JOURNAL   DNA Res. (2002) In press
REFERENCE   2  (bases 1 to 2593857)
  AUTHORS   Kaneko,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2002) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; 2-6-7
            Kazusa-kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/cyano/Thermo/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BA000039.
FEATURES             Location/Qualifiers
     source          1..2593857
                     /organism="Thermosynechococcus elongatus BP-1"
                     /db_xref="taxon:197221"

....

  gene            complement(486379..487815)
                     /gene="ycf24"
     CDS             complement(486379..487815)
                     /gene="ycf24"
                     /note="ORF_ID:tll0490"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter subunit"
                     /protein_id="NP_681280.1"
                     /db_xref="GI:22298033"
                     /translation="MSATVQSLVNQPYKYGFVTPIETETIPKGLNEDIIRLISAKKNE
                     PEFMLEFRLRAYRQWLKMSEPQWPRVSYPPINYQDIVYYSAPKQKEKLKSLDEVDPVL
                     LETFEKLGIPLSEQKRLTNVAVDAIFDSVSVATTFREELAKQGIIFCSISEALQDYPE
                     LVQKYLGSVVPIGDNFYAALNSAVFSDGSFVYVPKNTRCPMELSTYFRINNGESGQFE
                     RTLIIADAGSYVSYLEGCTAPMFDTNQLHAAVVELVALDNAEIKYSTVQNWYAGDENG
                     KGGIYNFVTKRGLCLGRNSKISWTQVETGSAITWKYPSCVLVGDNSVGEFYSVALTNH
                     YQQADTGTKMIHIGKNTRSRIVSKGISAGHSQNSYRGLVKIGPKATGARNYSQCDSML
                     IGDTAAANTFPYIQVQNPTAQVEHEASTSKIGEDQLFYFAQRGISAEDAVSMMISGFC
                     RDVFNQLPMEFAVEADRLLSLKLEGSVG"
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
3:  NC_003272. Nostoc sp. PCC 71...[gi:17227497]

Links


LOCUS       NC_003272            6413771 bp    DNA     circular BCT 10-DEC-2002
DEFINITION  Nostoc sp. PCC 7120 complete genome.
ACCESSION   NC_003272
VERSION     NC_003272.1  GI:17227497
KEYWORDS    .
SOURCE      Nostoc sp. PCC 7120
  ORGANISM  Nostoc sp. PCC 7120
            Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
REFERENCE   1
  AUTHORS   Kaneko,T., Nakamura,Y., Wolk,C.P., Kuritz,T., Sasamoto,S.,
            Watanabe,A., Iriguchi,M., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kohara,M., Matsumoto,M., Matsuno,A., Muraki,A.,
            Nakazaki,N., Shimpo,S., Sugimoto,M., Takazawa,M., Yamada,M.,
            Yasuda,M. and Tabata,S.
  TITLE     Complete genomic sequence of the filamentous nitrogen-fixing
            cyanobacterium Anabaena sp. strain PCC 7120
  JOURNAL   DNA Res. 8 (5), 205-213 (2001)
  MEDLINE   21595285
   PUBMED   11759840
REFERENCE   2  (bases 1 to 6413771)
  AUTHORS   Kaneko,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/cyanobase/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BA000019.
            COMPLETENESS: full length.
FEATURES             Location/Qualifiers
     source          1..6413771
                     /organism="Nostoc sp. PCC 7120"
                     /db_xref="taxon:103690"

....
 gene            2991689..2993128
                     /gene="alr2492"
     CDS             2991689..2993128
                     /gene="alr2492"
                     /note="probable ABC transporter membrane protein (ycf24)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="NP_486532.1"
                     /db_xref="GI:17229984"
                     /translation="MSATVKTLVNQPYKYGFVTDIEADTIPRGLDEDVVRLISTKKNE
                     PEFMLEFRLRAFRQWQKMTEPTWPSVKYPPIDYQNIIYYSAPKQKKAKLNSLDEVDPT
                     LIETFEKLGIPLSEQKRLANVAVDAIFDSVSVATTFKEKLAKDGVIFCSISEALQEHP
                     ELIKKYLGSVVPIADNYFAALNAAVFSDGSFVYIPKGVKCPMELSTYFRINSGDTGQF
                     ERTLIVAEEGSYVSYLEGCTAPMYDSNQLHAAVVELVALDNAEIKYSTVQNWYAGDAN
                     GKGGIYNFVTKRGLCQGVNSKISWTQVETGSAITWKYPSCVLVGDNSVGEFYSVALTN
                     NMQQADTGTKMIHIGKNTRSTIISKGISAGQSSNSYRGLVKINPTAKGARNYSQCDSM
                     LIGDNAHANTFPYIQVQNNTGKVEHEASTSKIGEDQLFFFAQRGISSEDAISMMISGF
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

CKDVFNQLPMEFAVEADKLLSLKLEGSVG"

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
4: NC_000911. Synechocystis sp....[gi:16329170]

Links


LOCUS       NC_000911            3573470 bp    DNA     circular BCT 09-DEC-2002
DEFINITION  Synechocystis sp. PCC 6803, complete genome.
ACCESSION   NC_000911
VERSION     NC_000911.1  GI:16329170
KEYWORDS    .
SOURCE      Synechocystis sp. PCC 6803
  ORGANISM  Synechocystis sp. PCC 6803
            Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
REFERENCE   1  (bases 1 to 3573470)
  AUTHORS   Kaneko,T., Tanaka,A., Sato,S., Kotani,H., Sazuka,T., Miyajima,N.,
            Sugiura,M. and Tabata,S.
  TITLE     Sequence analysis of the genome of the unicellular cyanobacterium
            Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
            region from map positions 64% to 92% of the genome
  JOURNAL   DNA Res. 2 (4), 153-166 (1995)
  MEDLINE   96127529
   PUBMED   8590279
REFERENCE   2  (bases 1 to 3573470)
  AUTHORS   Kaneko,T., Sato,S., Kotani,H., Tanaka,A., Asamizu,E., Nakamura,Y.,
            Miyajima,N., Hirosawa,M., Sugiura,M., Sasamoto,S., Kimura,T.,
            Hosouchi,T., Matsuno,A., Muraki,A., Nakazaki,N., Naruo,K.,
            Okumura,S., Shimpo,S., Takeuchi,C., Wada,T., Watanabe,A.,
            Yamada,M., Yasuda,M. and Tabata,S.
  TITLE     Sequence analysis of the genome of the unicellular
cyanobacteriumSynechocystis sp. strain PCC6803. II. Sequence determination of the
            entire genome and assignment of potential protein-coding regions
  JOURNAL   DNA Res. 3 (3), 109-136 (1996)
  MEDLINE   97061201
   PUBMED   8905231
REFERENCE   3  (bases 1 to 3573470)
  AUTHORS   Tabata,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1996) Satoshi Tabata, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:tabata@kazusa.or.jp, URL:http://www.kazusa.or.jp/cyano/,
            Tel:81-438-52-3933(ex.2330), Fax:81-438-52-3934)
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BA000022.
            COMPLETENESS: full length.
FEATURES             Location/Qualifiers


....

 gene            2871674..2873116
                 /gene="ycf24"
                 /note="locus_tag: slr0074"
     CDS         2871674..2873116
                 /gene="ycf24"
                 /codon_start=1
                 /transl_table=11
                 /product="ABC transporter subunit"
```

APPENDIX D

Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

/protein_id="NP_442472.1"
/db_xref="GI:16331744"
/translation="MSSTTVKNLVNQPYKYGFVTNIEADAIPRGLSEDVVRLISAKKN
EPEFMLDFRLRAYRHWLTMAEPTWPAVHYPPIDYQDIIYYSAPKQSKKKLESLDEVDP
ALLETFEKLGIPLSEQKRLSNVAVDAIFDSVSIGTTFKEKLAEDGVIFCSISEALQEH
PDLVQKYLGSVVPTADNFFAALNSAVFSDGSFVFIPKGVKCPMELSTYFRINNGDTGQ
FERTLIIAEEGASVSYLEGCTAPMYDTNQLHAAVVELVALDNADIKYSTVQNWYAGDE
NGKGGIYNFVTKRGLCKGVNSKISWTQVETGSAITWKYPSCVLVGDNSVGEFYSIALT
NNKQQADTGTKMIHIGKNTRSIIISKGISAGNSANSYRGLVKMGPKAQGARNYSQCDS
MLIGDRAAANTFPYIQVDNNTAKVEHEASTSKIGEDQLFYFAQRGISEEDAVSMLVSG
FCKDVLNELPMEFAAEADKLLSLKLEGTVG"

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
5: NC_000916. Methanobacterium ...[gi:15678031]

Links


LOCUS       NC_000916            1751377 bp    DNA     circular BCT 10-DEC-2002
DEFINITION  Methanobacterium thermoautotrophicum str. Delta H complete genome.
ACCESSION   NC_000916
VERSION     NC_000916.1  GI:15678031
KEYWORDS    .
SOURCE      Methanothermobacter thermautotrophicus str. Delta H
            (Methanobacterium thermoautotrophicum str. deltaH)
  ORGANISM  Methanothermobacter thermautotrophicus str. Delta H
            Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
            Methanobacteriaceae; Methanothermobacter.
REFERENCE   1  (bases 1 to 1751377)
  AUTHORS   Smith,D.R., Doucette-Stamm,L.A., Deloughery,C., Lee,H.-M.,
            Dubois,J., Aldredge,T., Bashirzadeh,R., Blakely,D., Cook,R.,
            Gilbert,K., Harrison,D., Hoang,L., Keagle,P., Lumm,W., Pothier,B.,
            Qiu,D., Spadafora,R., Vicare,R., Wang,Y., Wierzbowski,J.,
            Gibson,R., Jiwani,N., Caruso,A., Bush,D., Safer,H., Patwell,D.,
            Prabhakar,S., McDougall,S., Shimer,G., Goyal,A., Pietrovski,S.,
            Church,G.M., Daniels,C.J., Mao,J.-i., Rice,P., Nolling,J. and
            Reeve,J.N.
  TITLE     Complete genome sequence of Methanobacterium thermoautotrophicum
            deltaH: functional analysis and comparative genomics
  JOURNAL   J. Bacteriol. 179 (22), 7135-7155 (1997)
  MEDLINE   98037514
   PUBMED   9371463
REFERENCE   2  (bases 1 to 1751377)
  AUTHORS   NCBI Microbial Genomes Annotation Project.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-2001) National Center for Biotechnology
            Information, NIH, Bethesda, MD 20894, USA
FEATURES             Location/Qualifiers
     source          1..1751377
                     /organism="Methanothermobacter thermautotrophicus str.
                     Delta H"
                     /strain="Delta H"
                     /db_xref="taxon:187420"
                     /clone="MTH"
                     /note="synonym: Methanobacterium thermoautotrophicum str.
                     deltaH"

....

 gene            1060742..1061974
                     /gene="MTH1150"
     CDS             1060742..1061974
                     /gene="MTH1150"
                     /note="Function Code:12.02 - Cell Processes, Transport of
                     carbohydrates organic acids alcohols and lipids ; similar
                     to, gp:GI:g726070 LN:MTU19364, p()=1.1E-203, pid=93%"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter subunit Ycf24"
                     /protein_id="NP_276278.1"
                     /db_xref="GI:15679161"
                     /db_xref="CDD:pfam01458"
```

APPENDIX D

Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

/db_xref="COG:COG0719"
/translation="MLRDTLKKAEKAREKKALYGEDIDLEKFIKEEAGEHEEVTRAKE
VPKEVQETLLRVGVDPEERERAGTFIQVDQSGICTTCASESIEIMGMNVALDKYSWLK
DYMWKAVAVDTDKYTATTALREAEGEMGGYFIRSKPGAREVFPLQACMFIGDERVMQT
AHNIVIAEENSELHIITGCATGEDVSSALHVGVSEFYLKKGARITFTMVHNWAEQVEV
RPRTGIMVGDDATYINNYILTSPVKSIQSYPTAYCTGENSRVVFQSILGGQKDSVLDM
GSRVILEGRGSSAEMVSRAVSKDSSQIYSRGHLAGRVPEVKGHLECHGLVLSDDSMIY
AVPELEGSATELEMSHEAAVGKIAEEEVMYLTSRGLTEEEAASMIVRGFLSMDITGLP
PELAAETKRMLDMSLKGM"

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

6: NC_001799. Toxoplasma gondii...[gi:11496534]

Links

```
LOCUS       NC_001799              34996 bp    DNA     circular INV 19-SEP-2002
DEFINITION  Toxoplasma gondii apicoplast, complete genome.
ACCESSION   NC_001799
VERSION     NC_001799.1  GI:11496534
KEYWORDS    .
SOURCE      apicoplast Toxoplasma gondii
  ORGANISM  Toxoplasma gondii
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Sarcocystidae; Toxoplasma.
REFERENCE   1  (bases 4337 to 4925)
  AUTHORS   Beckers,C.J., Roos,D.S., Donald,R.G., Luft,B.J., Schwab,J.C.,
            Cao,Y. and Joiner,K.A.
  TITLE     Inhibition of cytoplasmic and organellar protein synthesis in
            Toxoplasma gondii. Implications for the target of macrolide
            antibiotics
  JOURNAL   Journal of Clin. Invest. 95, 367-376 (1995)
REFERENCE   2  (bases 13791 to 14996)
  AUTHORS   Kohler,S., Delwiche,C.F., Denny,P.W., Tilney,L.G., Webster,P.,
            Wilson,R.J., Palmer,J.D. and Roos,D.S.
  TITLE     A plastid of probable green algal origin in Apicomplexan parasites
  JOURNAL   Science 275 (5305), 1485-1489 (1997)
  MEDLINE   97197911
   PUBMED   9045615
REFERENCE   3  (bases 13218 to 15870; 28669 to 34697)
  AUTHORS   Denny,P., Preiser,P., Williamson,D. and Wilson,I.
  TITLE     Evidence for a single origin of the 35 kb plastid DNA in
            Apicomplexans
  JOURNAL   Protist 149, 51-59 (1998)
REFERENCE   4  (bases 1 to 34996)
  AUTHORS   Kissinger,J.C., Donald,R.G., Moulton,A.L., Gutell,R., Aiello,D.P.,
            Lang-Unnasch,N. and Roos,D.S.
  TITLE     Mapping, cloning, and complete sequence annotation of the 35-kb
            plastid genome of Toxoplasma gondii
  JOURNAL   Unpublished
REFERENCE   5  (bases 1 to 34996)
  AUTHORS   Kissinger,J.C., Donald,R.G., Moulton,A.L., Aiello,D.P.,
            Lang-Unnasch,N. and Roos,D.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Biology, University of Pennsylvania, 415 S.
            University Ave, Philadelphia, PA 19104, USA
REFERENCE   6  (bases 1 to 34996)
  AUTHORS   Kissinger,J.C., Donald,R.G., Moulton,A.L., Aiello,D.P.,
            Lang-Unnasch,N. and Roos,D.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1999) Biology, University of Pennsylvania, 415 S.
            University Ave, Philadelphia, PA 19104, USA
  REMARK    Sequence update by submitter
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from U87145.
            The T. gondii apicoplast genome exhbits non-standard codon usage:
            33 in-frame UGA codons interrupt 17 of the 28 predicted coding
            regions, and are presumed to encode tryptophan.  In addition, two
            genes (rps8 and rpoC2) contain in-frame UAA or UAG stop codons
            which are currently being investigated further.
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
...
CDS             complement(28289..29686)
                    /note="ycf24 homolog; in frame UGA codon predicted to
                    encode tryptophan; similar to Plasmodium falciparum
                    plastid genome ORF470"
                    /codon_start=1
                    /transl_except=(pos:complement(28916..28918),aa:Trp)
                    /transl_table=11
                    /product="ABC transporter"
                    /protein_id="NP_044570.1"
                    /db_xref="GI:11496556"
                    /translation="MKLYKYLYNKYNNNTDLFNTVRLIGGLNINMVNKLIFKQDNFIF
                    LYIFRLNALSILNKFKQPDWCFYELPEFAFDDISYYSIPLNVYTNKNKYKSILSKLGL
                    ELKFSENLILDVIFDSVLLLNLTTFFLIKMGLFFLSFFQSIIFYPYLIFSYLGSIVSN
                    TDNFFLTINSIIFNEGSFCFVMKDLNSNINLTTYFRTHSENFAQFERTLIVLSENSKL
                    IYFEGCSAPMFLESQLHIAIVELFIKTKANLKYSTIQNWYRGNQLGEGGLYNFTTKRG
                    FCMDKSFLNWIQIEIGSVITWKYPSTYLIGNKSFSNFFSLAMLSDYQVSDTGTKMLHI
                    GKNTKSFILSKSLSFNFSFYTYRGLVTIFKTALNSYNYTECNSLLIGCNAFTATIPYT
                    IINNFSAYINQEATISKLELDFLFFLLHRGLNLKSTLMILIYGYCYNICSKISFELEL
                    EVPLLIVARAQKLFY"
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

7: NC_001713. Odontella sinensi...[gi:11467432]

Links

```
LOCUS       NC_001713             119704 bp    DNA     circular PLN 19-SEP-2002
DEFINITION  Odontella sinensis chloroplast, complete genome.
ACCESSION   NC_001713
VERSION     NC_001713.1  GI:11467432
KEYWORDS    .
SOURCE      chloroplast Odontella sinensis
  ORGANISM  Odontella sinensis
            Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
            Biddulphiophycidae; Eupodiscales; Eupodiscaceae; Odontella.
REFERENCE   1  (bases 1 to 119704)
  AUTHORS   Kowallik,K.V., Stoebe,B., Schaffran,I., Kroth-Pancic,P. and
            Freier,U.
  TITLE     The Chloroplast Genome of a chlorophyll a+c- containing Alga,
            Odontella sinensis
  JOURNAL   Plant Mol. Biol. Rep. 13, 336-342 (1995)
REFERENCE   2  (bases 1 to 119704)
  AUTHORS   Kowallik,K.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-NOV-1995) Heinrich-Heine Universitaet Duesseldorf,
            Universitaetsstr. 1 Geb.26 13/02/46, Duesseldorf D-40225, Germany
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from Z67753.
FEATURES             Location/Qualifiers
     source          1..119704
                     /organism="Odontella sinensis"
                     /organelle="plastid:chloroplast"
                     /db_xref="taxon:2839"

....
gene            69086..70546
                     /gene="ycf24"
     CDS             69086..70546
                     /gene="ycf24"
                     /note="ORF486"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter"
                     /protein_id="NP_043655.1"
                     /db_xref="GI:11467509"
                     /db_xref="SWISS-PROT:P49530"
                     /translation="MTNKSNKILNTNITKLVNQPYKYGFSTVIEKDIIEKGLNEDVIC
                     LISKKKNEPKFLLEFRLKAFKKWKEMKCPDWAQIKFSEIDYQDIIYYSAPKVKKKLNS
                     LDEVDPELLKTFEKLGISLTEQKRLANVAIDAVFDSVSIATTFKEELAECGVIFSSIS
                     EAIQEYPELIEKYLGSVVPIGDNYFSALNSAVFTDGSFCYIPKDTICPLELSTYFRIN
                     DQKSGQFERTLIVAEKNSQVSYLEGCTAPQYDSNQLHAAVVELVALENADIKYSTVQN
                     WYAGNNYGEGGIYNFVTKRGLCAGSNSKISWTQVETGSNITWKYPSCLLVGDKAKGEF
                     YSVALTNNYQQADTGSKMIHVGKNTRSRIVSKGISAGNSKNTYRGLVNISNKAIGARN
                     YSQCDSLLIGNLSNANTFPFISVQNPTAKIEHEASTSKIGEEQIFYFLQRGIPIEKGV
                     ELMISGFCQEVFTELPLEFAAEADRLLTLKLEGSVG"
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
8: NC_001675. Cyanophora parado...[gi:11467282]

Links


LOCUS       NC_001675              135599 bp    DNA     circular PLN 19-SEP-2002
DEFINITION  Cyanophora paradoxa cyanelle, complete genome.
ACCESSION   NC_001675
VERSION     NC_001675.1  GI:11467282
KEYWORDS    .
SOURCE      cyanelle Cyanophora paradoxa
  ORGANISM  Cyanophora paradoxa
            Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
REFERENCE   1  (bases 1 to 135599)
  AUTHORS   Stirewalt,V.L., Michalowski,C.B., Luffelhardt,W., Bohnert,H.J. and
            Bryant,D.A.
  TITLE     Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 135599)
  AUTHORS   Bryant,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1995) Donald A. Bryant, Biochemistry and
            Molecular Biology, The Pennsylvania State University, S-234 Frear
            Bldg., University Park, PA 16802, USA
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from U30821.
FEATURES             Location/Qualifiers
     source          1..135599
                     /organism="Cyanophora paradoxa"
                     /organelle="plastid:cyanelle"
                     /strain="Pringsheim LB555"
                     /db_xref="taxon:2762"
...

 gene            complement(67893..69353)
                     /gene="ycf24"
     CDS             complement(67893..69353)
                     /gene="ycf24"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter"
                     /protein_id="NP_043219.1"
                     /db_xref="GI:11467362"
                     /translation="MVNTQSPKNSGLENLVNQPYKYGLPLIFEIETISKGLTEETIRL
                     ISEKKNEPQFMLEFRLQAYRKWLEMSNEPEWAHLNYPKINYQDMVYYSAPKQKKKLQS
                     LDEVDPTLLETFEKLGIPLTEQKRLANVAVDAIFDSVSVATTFKEELAKEGVIFCPIS
                     EAVQKYPDLIKKYLGSVVSTSDNYFSCLNAAVFSDGSFCYIPKNVRCPLELSTYFRIN
                     NGESGQFERTLIVADEGSYVSYLEGCTAPQFDTNQLHAAVVELVALDNAEIKYSTVQN
                     WYAGDENGKGGIYNFVTKRGLCAGKNSKISWTQVETGSAITWKYPSCVLLGDNSIGEF
                     YSVALTNRYQQADTGTKMIHIGKNTRSRIISKGISAGHSQNSYRGLVKIGPKAVGARN
                     YSQCDSLLIGDNSQANTFPHLQIKNPTAKVEHEASTSKIGEEQIFYFLQRGINAEEAI
                     SLIISGFCREVFNNLPMEFALEADKLLGLKLEGSVG"
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

9: NC_000925. Porphyra purpurea...[gi:11465652]

Links

```
LOCUS       NC_000925             191028 bp    DNA     circular PLN 27-AUG-2002
DEFINITION  Porphyra purpurea chloroplast, complete genome.
ACCESSION   NC_000925
VERSION     NC_000925.1  GI:11465652
KEYWORDS    .
SOURCE      chloroplast Porphyra purpurea
  ORGANISM  Porphyra purpurea
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 191028)
  AUTHORS   Reith,M.E. and Munholland,J.
  TITLE     Complete nucleotide sequence of the Porphyra purpurea chloroplast
            genome
  JOURNAL   Plant Mol. Biol. Rep. 13 (4), 333-335 (1995)
REFERENCE   2  (bases 1 to 191028)
  AUTHORS   Reith,M.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-1995) Michael E. Reith, Marine Biology Section,
            NRC Institute for Marine Biosciences, 1411 Oxford Street, Halifax,
            Nova Scotia B3H 3Z1, Canada
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from U38804.
FEATURES             Location/Qualifiers
     source          1..191028
                     /organism="Porphyra purpurea"
                     /organelle="plastid:chloroplast"
                     /strain="Avonport"
                     /db_xref="taxon:2787"

...

gene            40948..42411
                     /gene="ycf24"
     CDS             40948..42411
                     /gene="ycf24"
                     /note="hypothetical chloroplast ORF 24"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter"
                     /protein_id="NP_053850.1"
                     /db_xref="GI:11465706"
                     /translation="MVNTQNQISQTSDLDYIVNQPYKYGFTTSVESEQFPRGISREVV
                     KLISKKKNEPEYLLNFRLKAYEKWTKMKNPKWAHLKHPNIDFNSIIYYAVPKLKKELN
                     SLDEVDPEILDTFNKLGISLNEQKRLSNVAVDAVFDSVSIATTFKKELAEAGVIFCSI
                     SEAIRNYPDLIQKYLGTVVPSGDNYFAALNSAVFSDGSFCYIPPDTVCPLELSTYFRI
                     NNEESGQFERTLIVADRGSKVSYLEGCTAPQYDTNQLHAAIVELIALDDAEIKYSTVQ
                     NWYAGNKDGKGGIYNFVTKRGLCSGKNSKISWTQVETGSAITWKYPGCILAGDNSQGE
                     FYSVALTNNYQEADTGTKMIHIGNNTKSKIISKGISAGKSKNSYRGLVKIGPQSFNSR
                     NYSQCDSLLIGQSSQANTFPYIQVQNPTAKVEHEASTSKISEDQIFYFLQRGINLEES
                     VSLMISGFCKDVFNELPMEFAVEADRLLSLKLEGTVG"
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

10: NC_000926. Guillardia theta ...[gi:11467607]

Links

```
LOCUS       NC_000926             121524 bp    DNA     circular PLN 22-AUG-2002
DEFINITION  Guillardia theta chloroplast, complete genome.
ACCESSION   NC_000926
VERSION     NC_000926.1  GI:11467607
KEYWORDS    .
SOURCE      chloroplast Guillardia theta
  ORGANISM  Guillardia theta
            Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
REFERENCE   1  (bases 47701 to 48415)
  AUTHORS   Douglas,S.E. and Durnford,D.G.
  TITLE     The small subunit of ribulose-1,5-bisphosphate carboxylase is
            plastid-encoded in the chlorophyll c-containing alga Cryptomonas
            phi
  JOURNAL   Plant Mol. Biol. 13 (1), 13-20 (1989)
  MEDLINE   93357429
   PUBMED   2562756
REFERENCE   2  (bases 18535 to 19351)
  AUTHORS   Douglas,S.E. and Durnford,D.G.
  TITLE     Sequence analysis of the plastid rDNA spacer region of the
            chlorophyll c-containing alga Cryptomonas phi
  JOURNAL   DNA Seq. 1 (1), 55-62 (1990)
  MEDLINE   92119320
   PUBMED   2132959
REFERENCE   3  (bases 43739 to 44938)
  AUTHORS   Douglas,S.E. and Durnford,D.G.
  TITLE     Nucleotide sequence of the genes for ribosomal protein S4 and
            tRNA(Arg) from the chlorophyll c-containing alga Cryptomonas phi
  JOURNAL   Nucleic Acids Res. 18 (7), 1903 (1990)
  MEDLINE   90245597
   PUBMED   2336372
REFERENCE   4  (bases 34539 to 35380)
  AUTHORS   Reith,M. and Douglas,S.
  TITLE     Localization of beta-phycoerythrin to the thylakoid lumen of
            Cryptomonas phi does not involve a signal peptide
  JOURNAL   Plant Mol. Biol. 15 (4), 585-592 (1990)
  MEDLINE   91338697
   PUBMED   2102376
REFERENCE   5  (bases 45872 to 47981)
  AUTHORS   Douglas,S.E., Durnford,D.G. and Morden,C.W.
  TITLE     Nucleotide sequence of the gene for the large subunit of
            ribulose-1,5-bisphosphate carboxylase/oxygenase from the
            chlorophyll c-containing Alga Cryptomonas F: evidence supporting
            the polyphyletic origin of plastids
  JOURNAL   J. Phycol. 26, 500-508 (1990)
REFERENCE   6  (bases 110917 to 113854)
  AUTHORS   Douglas,S.E.
  TITLE     Unusual organization of a ribosomal protein operon in the plastid
            genome of Cryptomonas phi: evolutionary considerations
  JOURNAL   Curr. Genet. 19 (4), 289-294 (1991)
  MEDLINE   91330343
   PUBMED   1868578
REFERENCE   7  (bases 40675 to 42376)
  AUTHORS   Douglas,S.E. and Turner,S.
  TITLE     Molecular evidence for the origin of plastids from a
```

APPENDIX D
Excerpts from 26 *ycf24* gene product and gene sequences retrieved from NCBI database

```
            cyanobacterium-like ancestor
  JOURNAL   J. Mol. Evol. 33 (3), 267-273 (1991)
  MEDLINE   92099311
   PUBMED   1757997
REFERENCE   8  (bases 96129 to 98906)
  AUTHORS   Wang,S.L. and Liu,X.Q.
  TITLE     The plastid genome of Cryptomonas phi encodes an hsp70-like
            protein, a histone-like protein, and an acyl carrier protein
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88 (23), 10783-10787 (1991)
  MEDLINE   92073372
   PUBMED   1961745
REFERENCE   9  (bases 106789 to 108216)
  AUTHORS   Douglas,S.E.
  TITLE     A secY homologue is found in the plastid genome of Cryptomonas phi
  JOURNAL   FEBS Lett. 298 (1), 93-96 (1992)
  MEDLINE   92183838
   PUBMED   1544427
REFERENCE   10 (bases 42198 to 44153)
  AUTHORS   Douglas,S.E. and Reith,M.E.
  TITLE     A bchI homolog, encoding a subunit of Mg chelatase, is located on
            the plastid genomes of red and cryptomonad algae
  JOURNAL   J. Mar. Biotechnol. 1, 135-141 (1993)
REFERENCE   11 (bases 82327 to 84479)
  AUTHORS   Douglas,S.E. and Murphy,C.A.
  TITLE     Structural, transcriptional and phylogenetic analyses of the atpB
            gene cluster from the plastid of Cryptomonas F (Cryptophyceae)
  JOURNAL   J. Phycol. 30, 329-340 (1994)
REFERENCE   12 (bases 98901 to 114602)
  AUTHORS   Wang,S.L., Liu,X.Q. and Douglas,S.E.
  TITLE     The large ribosomal protein gene cluster of a cryptomonad plastid:
            gene organization, sequence and evolutionary implications
  JOURNAL   Biochem. Mol. Biol. Int. 41 (5), 1035-1044 (1997)
  MEDLINE   97283757
   PUBMED   9137835
REFERENCE   13 (bases 61067 to 68605)
  AUTHORS   Leitsch,C.E.W., Kowallik,K.V. and Douglas,S.E.
  TITLE     The atpA gene cluster of a cryptomonad, Guillardia theta: A piece
            in the puzzle of chloroplast genome development
  JOURNAL   J. Phycol. (1998) In press
REFERENCE   14 (bases 1 to 121524)
  AUTHORS   Douglas,S.E. and Penny,S.L.
  TITLE     The plastid genome of the cryptophyte alga, Guillardia theta:
            complete sequence and conserved synteny groups confirm its common
            ancestry with red algae
  JOURNAL   J. Mol. Evol. 48 (2), 236-244 (1999)
  MEDLINE   99128221
   PUBMED   9929392
REFERENCE   15 (bases 1 to 121524)
  AUTHORS   Douglas,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-1998) Institute for Marine Biosciences, National
            Research Council, 1411 Oxford Street, Halifax, Nova Scotia B3H 3Z1,
            Canada
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from AF041468.
FEATURES             Location/Qualifiers
     source          1..121524
                     /organism="Guillardia theta"
                     /organelle="plastid:chloroplast"
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
                     /db_xref="taxon:55529"
                     /note="plastid"

...

gene            61067..62518
                     /gene="ycf24"
     CDS             61067..62518
                     /gene="ycf24"
                     /note="hypothetical chloroplast RF24"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter"
                     /protein_id="NP_050730.1"
                     /db_xref="GI:11467678"
                     /translation="MSDDLSKRSLRELVSQPYKYGFHTDIENEEFPKGLDEDIIKEIS
                     KLKCEPSYMLDFRLKSYILWKKMSLPEWACLTYLNINYQDIVYYSAPKNSTKLDSLED
                     VDKKILETFDKLGIPLNEQKKLANVAVDAIFDSVSVGTTFKQELSNVGVLFCPLSEAT
                     NKYSTLVEKYLGSVVPIGDNYFAALNSAVFSEGSFCYIPPNVKCPLELSTYFRINNEN
                     SGQFERTLIIADFNSYVSYLEGCTAPMYDKNQLHAAVVELIALENAEIRYSTVQNWYS
                     GDTNGKGGIYNFVTKRGLCAGKSSKISWTQVETGSAITWKYPSCILVGEDSVGEFYSV
                     ALTNNYQQADTGTKMIHVGRGSKSRIISKGISAGYSKNTYRGQVKININALGSINNSQ
                     CDSMLIGPYSQANTYPYIQVSNAMSRVEHEASTSKIEEEQLFYFLQRGISVEQAISLL
                     ISGFCRDVFVKLPMEFAVEADKLLSVKLEGTVG"
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

11:NC_001840. Cyanidium caldari...[gi:11465393]

Links

```
LOCUS       NC_001840              164921 bp    DNA     circular PLN 22-AUG-2002
DEFINITION  Cyanidium caldarium chloroplast, complete genome.
ACCESSION   NC_001840
VERSION     NC_001840.1  GI:11465393
KEYWORDS    .
SOURCE      chloroplast Cyanidium caldarium
  ORGANISM  Cyanidium caldarium
            Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales;
            Porphyridiaceae; Cyanidium.
REFERENCE   1  (bases 130696 to 132364)
  AUTHORS   Vogel,H., Fischer,S. and Valentin,K.
  TITLE     A model for the evolution of the plastid sec apparatus inferred
            from secY gene phylogeny
  JOURNAL   Plant Mol. Biol. 32 (4), 685-692 (1996)
  MEDLINE   97134960
   PUBMED   8980520
REFERENCE   2  (bases 1 to 164921)
  AUTHORS   Glockner,G., Rosenthal,A. and Valentin,K.
  TITLE     The structure and gene repertoire of an ancient red algal plastid
            genome
  JOURNAL   J. Mol. Evol. 51 (4), 382-390 (2000)
  MEDLINE   20496959
   PUBMED   11040290
REFERENCE   3  (bases 46857 to 47851)
  AUTHORS   Valentin,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-1996) Institute for Plant Physiology, Justus
            Liebig University, Heinrich Buff Ring 58-62, Giessen 35392, Germany
REFERENCE   4  (bases 28701 to 75580)
  AUTHORS   Gloeckner,G., Rosenthal,A. and Valentin,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Department of Genome Analysis, IMB Jena,
            Beutenbergstr.11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 164921)
  AUTHORS   Gloeckner,G., Rosenthal,A. and Valentin,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1999) Genome Analysis, Institute for Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE   6  (bases 130696 to 132364)
  AUTHORS   Vogel,H., Fischer,S. and Valentin,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1999) Institute for Plant Physiology, Justus
            Liebig University, Heinrich Buff Ring 58-62, Giessen 35392, Germany
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from AF022186.
FEATURES             Location/Qualifiers
     source          1..164921
                     /organism="Cyanidium caldarium"
                     /organelle="plastid:chloroplast"
                     /strain="RK1"
                     /db_xref="taxon:2771"
     repeat_region   join(164808..164921,1..78,462..644)
                     /note="repeat unit separated by stem loop"
                     /rpt_type=direct
```

APPENDIX D

Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
   stem_loop       339..477
                   /note="stem-loop separates direct repeat unit"
   rep_origin      580..780
                   /note="similiar to yeast mitochondrial origin of
                   replication region in GenBank Accession Number L36902"
...

gene            114386..115837
                   /gene="ycf24"
   CDS             114386..115837
                   /gene="ycf24"
                   /note="similar to ycf24 in Porphyra purpurea"
                   /codon_start=1
                   /transl_table=11
                   /product="ABC transporter"
                   /protein_id="NP_045146.1"
                   /db_xref="GI:11465463"
                   /translation="MIDRKKSSNIQNILNKPYKYGFSTEIQSEEFPKGINEEIIRLMS
                   HKKQEPDFILKFRLKAYQIWKKMQAPDWGHLHHNEINFNDVLCYASPKLEQGKNKAQT
                   ISEEILATFEKLGVPIKPNNKQPKIAVDAVFDSISFGTTLQKELKEQGIIFCSISEAI
                   KAYPNLIKKYLGSIVPAGDNYFAALNSAVFTDGSFCYIPKNIRCPVDLSTYFRINNKE
                   AGQFERTLIIADENSFVNYLEGCTAPQFDTNQLHAAVVELICFKNATINYSTVQNWYA
                   GNNKGEGGVYNFVTKRGLCQGENSKISWTQLETGSAITWKYPSCLLKGKRSTGEFFSV
                   TLTNNAQEADTGTKMLHFGRQSKSLVISKGISGGVSKNTYRGLVKISGSAIYSDNRSQ
                   CDSLLIGKGSESNTYPNLHVHNSLSKVEHEAFVSRIGEEQIFYFQQRGINIEEALNMI
                   VSGFCQDVCNKLPMEFALEANKLLNIKLEGSIG"
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

12: AP005370. Thermosynechococc...[gi:22294033]

Links

```
LOCUS       AP005370              299350 bp    DNA     linear   BCT 17-AUG-2002
DEFINITION  Thermosynechococcus elongatus BP-1 DNA, complete genome, section
            2/9.
ACCESSION   AP005370 BA000039
VERSION     AP005370.1  GI:22294033
KEYWORDS    .
SOURCE      Thermosynechococcus elongatus BP-1
  ORGANISM  Thermosynechococcus elongatus BP-1
            Bacteria; Cyanobacteria; Chroococcales; Thermosynechococcus.
REFERENCE   1
  AUTHORS   Nakamura,Y., Kaneko,T., Sato,S., Ikeuchi,M., Katoh,H., Sasamoto,S.,
            Watanabe,A., Iriguchi,M., Kawashima,K., Kimura,T., Kishida,Y.,
            Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A., Nakazaki,N.,
            Shimpo,S., Sugimoto,M., Takeuchi,C., Yamada,M. and Tabata,S.
  TITLE     Complete genome structure of the thermophilic cyanobacterium
            Thermosynechococcus elongatus BP-1
  JOURNAL   DNA Res. (2002) In press
REFERENCE   2  (bases 1 to 299350)
  AUTHORS   Kaneko,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2002) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; 2-6-7
            Kazusa-kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/cyano/Thermo/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1..299350
                     /organism="Thermosynechococcus elongatus BP-1"
                     /strain="BP-1"
                     /db_xref="taxon:197221"
                     /note="BP-1"
...
  gene            complement(185829..187265)
                     /gene="ycf24"
     CDS             complement(185829..187265)
                     /gene="ycf24"
                     /note="ORF_ID:tll0490"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter subunit"
                     /protein_id="BAC08042.1"
                     /db_xref="GI:22294211"
                     /translation="MSATVQSLVNQPYKYGFVTPIETETIPKGLNEDIIRLISAKKNE
                     PEFMLEFRLRAYRQWLKMSEPQWPRVSYPPINYQDIVYYSAPKQKEKLKSLDEVDPVL
                     LETFEKLGIPLSEQKRLTNVAVDAIFDSVSVATTFREELAKQGIIFCSISEALQDYPE
                     LVQKYLGSVVPIGDNFYAALNSAVFSDGSFVYVPKNTRCPMELSTYFRINNGESGQFE
                     RTLIIADAGSYVSYLEGCTAPMFDTNQLHAAVVELVALDNAEIKYSTVQNWYAGDENG
                     KGGIYNFVTKRGLCLGRNSKISWTQVETGSAITWKYPSCVLVGDNSVGEFYSVALTNH
                     YQQADTGTKMIHIGKNTRSRIVSKGISAGHSQNSYRGLVKIGPKATGARNYSQCDSML
                     IGDTAAANTFPYIQVQNPTAQVEHEASTSKIGEDQLFYFAQRGISAEDAVSMMISGFC
                     RDVFNQLPMEFAVEADRLLSLKLEGSVG"
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

13:AE000884. Methanobacterium ...[gi:2622242]

Links

```
LOCUS       AE000884               11204 bp    DNA     linear   BCT 19-JUN-2002
DEFINITION  Methanobacterium thermoautotrophicum from bases 1050856 to 1062059
            (section 90 of 148) of the complete genome.
ACCESSION   AE000884 AE000666
VERSION     AE000884.1  GI:2622242
KEYWORDS    .
SOURCE      Methanothermobacter thermautotrophicus str. Delta H
            (Methanobacterium thermoautotrophicum str. deltaH)
  ORGANISM  Methanothermobacter thermautotrophicus str. Delta H
            Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
            Methanobacteriaceae; Methanothermobacter.
REFERENCE   1  (bases 1 to 11204)
  AUTHORS   Smith,D.R., Doucette-Stamm,L.A., Deloughery,C., Lee,H., Dubois,J.,
            Aldredge,T., Bashirzadeh,R., Blakely,D., Cook,R., Gilbert,K.,
            Harrison,D., Hoang,L., Keagle,P., Lumm,W., Pothier,B., Qiu,D.,
            Spadafora,R., Vicaire,R., Wang,Y., Wierzbowski,J., Gibson,R.,
            Jiwani,N., Caruso,A., Bush,D. and Reeve,J.N.
  TITLE     Complete genome sequence of Methanobacterium thermoautotrophicum
            deltaH: functional analysis and comparative genomics
  JOURNAL   J. Bacteriol. 179 (22), 7135-7155 (1997)
  MEDLINE   98037514
   PUBMED   9371463
REFERENCE   2  (bases 1 to 11204)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-AUG-1997) Genomics and Technology Development, Genome
            Therapuetics Corporation, 100 Beaver Street, Waltham, MA
            02154-8448, USA
FEATURES             Location/Qualifiers
     source          1..11204
                     /organism="Methanothermobacter thermautotrophicus str.
                     Delta H"
                     /strain="Delta H"
                     /db_xref="taxon:187420"
                     /clone="MTH"
                     /note="synonym: Methanobacterium thermoautotrophicum str.
                     deltaH"
     ....

     gene            9887..11119
                     /gene="MTH1150"
     CDS             9887..11119
                     /gene="MTH1150"
                     /note="Function Code:12.02 - Cell Processes, Transport of
                     carbohydrates organic acids alcohols and lipids ; similar
                     to, gp:GI:g726070 LN:MTU19364, p()=1.1E-203, pid=93%"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter subunit Ycf24"
                     /protein_id="AAB85639.1"
                     /db_xref="GI:2622255"
                     /translation="MLRDTLKKAEKAREKKALYGEDIDLEKFIKEEAGEHEEVTRAKE
                     VPKEVQETLLRVGVDPEERERAGTFIQVDQSGICTTCASESIEIMGMNVALDKYSWLK
                     DYMWKAVAVDTDKYTATTALREAEGEMGGYFIRSKPGAREVFPLQACMFIGDERVMQT
```

APPENDIX D

Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

AHNIVIAEENSELHIITGCATGEDVSSALHVGVSEFYLKKGARITFTMVHNWAEQVEV
RPRTGIMVGDDATYINNYILTSPVKSIQSYPTAYCTGENSRVVFQSILGGQKDSVLDM
GSRVILEGRGSSAEMVSRAVSKDSSQIYSRGHLAGRVPEVKGHLECHGLVLSDDSMIY
AVPELEGSATELEMSHEAAVGKIAEEEVMYLTSRGLTEEEAASMIVRGFLSMDITGLP
PELAAETKRMLDMSLKGM"

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

14: AP003589. Nostoc sp. PCC 71...[gi:17131372]

Links

```
LOCUS       AP003589              341880 bp    DNA     linear   BCT 28-NOV-2001
DEFINITION  Nostoc sp. PCC 7120 DNA, complete genome, section 9/19.
ACCESSION   AP003589 BA000019
VERSION     AP003589.1  GI:17131372
KEYWORDS    .
SOURCE      Nostoc sp. PCC 7120
  ORGANISM  Nostoc sp. PCC 7120
            Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
REFERENCE   1
  AUTHORS   Kaneko,T., Nakamura,Y., Wolk,C.P., Kuritz,T., Sasamoto,S.,
            Watanabe,A., Iriguchi,M., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kohara,M., Matsumoto,M., Matsuno,A., Muraki,A.,
            Nakazaki,N., Shimpo,S., Sugimoto,M., Takazawa,M., Yamada,M.,
            Yasuda,M. and Tabata,S.
  TITLE     Complete genomic sequence of the filamentous nitrogen-fixing
            cyanobacterium Anabaena sp. strain PCC 7120
  JOURNAL   DNA Res. 8 (5), 205-213 (2001)
  MEDLINE   21595285
   PUBMED   11759840
REFERENCE   2  (bases 1 to 341880)
  AUTHORS   Kaneko,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/cyanobase/,
            Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
FEATURES             Location/Qualifiers
     source          1..341880
                     /organism="Nostoc sp. PCC 7120"
                     /db_xref="taxon:103690"
                     /note="synonym:Anabaena sp. PCC7120"
     ...
     gene            244219..245658
                     /gene="alr2492"
     CDS             244219..245658
                     /gene="alr2492"
                     /note="ORF_ID:alr2492
                     probable ABC transporter membrane protein (ycf24)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB74191.1"
                     /db_xref="GI:17131584"
                     /translation="MSATVKTLVNQPYKYGFVTDIEADTIPRGLDEDVVRLISTKKNE
                     PEFMLEFRLRAFRQWQKMTEPTWPSVKYPPIDYQNIIYYSAPKQKKAKLNSLDEVDPT
                     LIETFEKLGIPLSEQKRLANVAVDAIFDSVSVATTFKEKLAKDGVIFCSISEALQEHP
                     ELIKKYLGSVVPIADNYFAALNAAVFSDGSFVYIPKGVKCPMELSTYFRINSGDTGQF
                     ERTLIVAEEGSYVSYLEGCTAPMYDSNQLHAAVVELVALDNAEIKYSTVQNWYAGDAN
                     GKGGIYNFVTKRGLCQGVNSKISWTQVETGSAITWKYPSCVLVGDNSVGEFYSVALTN
                     NMQQADTGTKMIHIGKNTRSTIISKGISAGQSSNSYRGLVKINPTAKGARNYSQCDSM
                     LIGDNAHANTFPYIQVQNNTGKVEHEASTSKIGEDQLFFFAQRGISSEDAISMMISGF
                     CKDVFNQLPMEFAVEADKLLSLKLEGSVG"
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

15: BI437350. gc59d11.y1 Moss E...[gi:15262040]

Links

IDENTIFIERS

dbEST Id: 9279936
EST name: gc59d11.y1
GenBank Acc: BI437350
GenBank gi: 15262040

CLONE INFO
Clone Id: PEP_SOURCE_ID:PPN200721 (5')
Source: University of Leeds (UK) & Washington University in St. Louis (USA)
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

GGGAAGTCTTTAAGGAGCTTCCTCTGGAGTTTGCTGCTGAGGTCAATCAATTGATGAGCT
TGAAGCTGGAGGGAAGCGTTGGCTAATTTATTTGTACCCAAGGTGTCGTTTCAAGAGATG
TAGACGACGCTTGTTGTCAAATGCTTATGTCCTAGTTGGAGAATGCTTCGGTCCTGCTTA
TCAAGATGTCTCCTGCTTTGTTTAAAACCAAGGTTCGAGTTCATGTATTTTAACATTTTT
ATATTCACACGCCACCAATGTCGGAAGTGAGGGAAACTGATTGCGAAGAGGGAATAGATA
TATACAATTCATGCGGGTGCGCGGTGGATGAATTCAACTTTTCACATATGCCTCAGCTGT

Entry Created: Aug 21 2001
Last Updated: Aug 21 2001

COMMENTS

Libraries were constructed by Dr. Stavros Bashiardes as part of the Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and Washington Univ. in St. Louis (USA) DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)

PUTATIVE ID Assigned by submitter
SW:YC24_ODOSI P49530 HYPOTHETICAL 54.3 KD PROTEIN YCF24 ;

LIBRARY
Lib Name: Moss EST library PPN
Organism: Physcomitrella patens
Tissue type: protonemata: 7 day old tissue auxin treated
Lab host: DH10B
Vector: pBluescript SK-
R. Site 1: EcoRI
R. Site 2: XhoI
Description: Construction of the cDNA library was carried out using Stratagenes 'UniZAP - cDNA synthesis kit'. cDNA was constructed using an oligo dT primer/linker that contains a XhoI site within it. Following ds cDNA synthesis, EcoRI adapters were ligated to the blunt ends and sample was digested with XhoI. The result is cDNA with an EcoRI sticky end on one side and a XhoI sticky end on the other. This cDNA was ligated directionally in UniZAP arms. The vector is

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

designed containing the pBluescript sequence as well as lambda DNA and cDNA is cloned within this pBluescript sequence. The vector was then packaged using Gold gigapackaging extracts. Library was grown in XLIBlue MRF' cells and amplified. The library was excised by mass excision using Stratagens 'Mass excision kit' that uses exassist as a helper phage that releases the pBluescript sequence and circularises it as single stranded plasmids that are then packaged (by helper phage) and secreted out of the host cell as phagemids. SOLR cells were transformed with phagemids and the library was plated out on LB-amp plates to select for transformants. Approximately 1,000,000 colonies were grown and recovered. The double stranded plasmid library was recovered by using Quiagen Midi prep kit. 2 micro grams of each library were used to transform DH10B cells by electroporation.

SUBMITTER
Name: Ralph Quatrano
Lab: Leeds/Wash U Moss EST Project
Institution: Washington University School of Medicine
Address: 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
E-mail: est@watson.wustl.edu

CITATIONS
Title: Leeds/Wash U Moss EST Project
Authors: Quatrano,R., Bashiardes,S., Cove,D., Cuming,A., Knight,C., Clifton,S., Marra,M., Hillier,L., Pape,D., Martin,J., Wylie ,T., Underwood,K., Theising,B., Allen,M., Bowers,Y., Person ,B., Swaller,T., Steptoe,M., Gibbons,M., Harvey,N., Ritter ,E., Jackson,Y., McCann,R., Waterston,R., Wilson,R.
Year: 1999
Status: Unpublished

MAP DATA

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

16:1: Z67753. Odontella sinensi...[gi:1185127]

Links

LOCUS OSCHLPLXX 119704 bp DNA circular PLN 10-APR-2001
DEFINITION Odontella sinensis complete chloroplast genome.
ACCESSION Z67753
VERSION Z67753.1 GI:1185127
KEYWORDS 16S ribosomal RNA; 16S rRNA gene; 23S ribosomal RNA; 23S rRNA gene; 30S ribosomal protein S10; 30S ribosomal protein S11; 30S ribosomal protein S12; 30S ribosomal protein S13; 30S ribosomal protein S14; 30S ribosomal protein S16; 30S ribosomal protein S17; 30S ribosomal protein S18; 30S ribosomal protein S19; 30S ribosomal protein S2; 30S ribosomal protein S20; 30S ribosomal protein S3; 30S ribosomal protein S4; 30S ribosomal protein S5; 30S ribosomal protein S6; 30S ribosomal protein S7; 30S ribosomal protein S8; 30S ribosomal protein S9; 50S ribosomal protein L1; 50S ribosomal protein L11; 50S ribosomal protein L12; 50S ribosomal protein L13; 50S ribosomal protein L14; 50S ribosomal protein L16; 50S ribosomal protein L18; 50S ribosomal protein L19; 50S ribosomal protein L2; 50S ribosomal protein L20; 50S ribosomal protein L21; 50S ribosomal protein L22; 50S ribosomal protein L23; 50S ribosomal protein L24; 50S ribosomal protein L27; 50S ribosomal protein L29; 50S ribosomal protein L3; 50S ribosomal protein L31; 50S ribosomal protein L32; 50S ribosomal protein L33; 50S ribosomal protein L34; 50S ribosomal protein L35; 50S ribosomal protein L36; 50S ribosomal protein L4; 50S ribosomal protein L5; 50S ribosomal protein L6; 5S ribosomal RNA; 5S rRNA gene; acp gene; acyl carrier protein; apocytochrome f; ATP synthase CF1; ATP synthase CFO; atpA gene; atpB gene; atpD gene; atpE gene; atpF gene; atpG gene; atpH gene; atpI gene; cfxQ gene; chaperonin; chlI gene; clp protease; clpC gene; complete genome; CP43 chlorophyll apoprotein; CP47 chlorophyll apoprotein; cytochrome b559; cytochrome b6; cytochrome b6/f complex; cytochrome c550; D1 reaction-center protein; D2 reaction-center protein; DNA-replication helicase; dnaB gene; dnaK gene; elongation factor tu; Fe-S polypeptide; ferredoxin; ferredoxin-binding protein II; groEL gene; orf 179; ORF 312; ORF 644; ORF113; ORF128; ORF148; ORF181; ORF204; ORF25; ORF251; ORF263; ORF26a; ORF26b; ORF27; ORF29; ORF29a; ORF29b; ORF31; ORF319; ORF34; ORF355; ORF355 gene; ORF36; ORF380; ORF382; ORF41; ORF42; ORF44; ORF455; ORF46; ORF486; ORF497; ORF61; ORF64; ORF73; ORF74; ORF99; P700 apoprotein A1; P700 apoprotein A2; petA gene; petB gene; petD gene; petF gene; petG gene; petK gene; phosphoprotein; plastocyanin-binding; preprotein-translocase; protein I; protein J; protein K; protein L; protein M; protein N; protein T; protein W; protein X; psaA gene; psaB gene; psaC gene; psaD gene; psaE gene; psaF gene; psaI gene; psaJ gene; psaL gene; psaM gene; psbA gene; psbB gene; psbC gene; psbD gene; psbE gene; psbF gene; psbH gene; psbI gene; psbJ gene; psbK gene; psbL gene; psbN gene; psbT gene; psbV gene; psbW gene; psbX gene; PSI; PSII; rbcL gene; rbcR gene; rbcs gene; RNA polymerase; RNA polymerase b''-chain; RNA polymerase b'-chain; RNA polymerase b-chain; rpL1 gene; rpl11 gene; rpl12 gene; rpl13 gene; rpl14 gene; rpl16 gene; Rpl18 gene; rpL19 gene; rpl2 gene; rpl20 gene; rpl21 gene; rpl22 gene; rpl23 gene; rpl24 gene; rpl27 gene; rpL29 gene; rpL3 gene; rpL31 gene; rpl32 gene; rpl32' gene; rpl33 gene; rpL34 gene; rpl35 gene; rpl36 gene; rpL4 gene; rpl5 gene; rpl6 gene; rpoA gene; rpoB gene; rpoC1 gene; rpoC2 gene; rps10 gene; rps11 gene; rps12 gene; rps13 gene; rps14 gene; rps16 gene;

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
            rps17 gene; rps18 gene; rps19 gene; rps2 gene; rps20 gene; rps3
            gene; rps4 gene; RPS5 gene; RPS6 gene; rps7 gene; rps8 gene; rps9
            gene; rubisco large subunit; rubisco small subunit; secA gene; secY
            gene; thiG gene; trnA gene; tRNA-Ala; tRNA-Arg; tRNA-Asn; tRNA-Asp;
            tRNA-Cys; tRNA-fMet; tRNA-Gln; tRNA-Glu; tRNA-Gly; tRNA-His;
            tRNA-Ile; tRNA-Leu; tRNA-Lys; tRNA-Met; tRNA-Phe; tRNA-Pro;
            tRNA-Ser; tRNA-Thr; tRNA-Trp; tRNA-Tyr; tRNA-Val; trnC gene; trnD
            gene; trnE gene; trnF gene; trnfM gene; trnG gene; trnH gene; trnI
            gene; trnK gene; trnL gene; trnM gene; trnN gene; trnP gene; trnQ
            gene; trnR gene; trnS gene; trnT gene; trnV gene; trnW gene; trnY
            gene; trsE gene; tufA gene; ycf12 gene; ycf16 gene; ycf24 gene;
            ycf25 gene; ycf3 gene; ycf30 gene; ycf31 gene; ycf32 gene; ycf33
            gene; ycf35 gene; ycf39 gene; ycf4 gene; ycf40 gene; ycf41 gene;
            ycf42 gene; ycf43 gene; ycf44 gene; ycf45 gene; ycf46 gene; ycf47
            gene; ycf5 gene; ycf6 gene; ycf7 gene; ycf9 gene.
SOURCE      chloroplast Odontella sinensis
  ORGANISM  Odontella sinensis
            Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
            Biddulphiophycidae; Eupodiscales; Eupodiscaceae; Odontella.
REFERENCE   1
  AUTHORS   Kowallik,K.V., Stoebe,B., Schaffran,I., Kroth-Pancic,P. and
            Freier,U.
  TITLE     The Chloroplast Genome of a chlorophyll a+c- containing Alga,
            Odontella sinensis
  JOURNAL   Plant Mol. Biol. Rep. 13, 336-342 (1995)
REFERENCE   2  (bases 1 to 119704)
  AUTHORS   Kowallik,K.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-NOV-1995) K.V. Kowallik, Heinrich-Heine Universitaet
            Duesseldorf, Universitaetsstr. 1 Geb.26 13/02/46, D- 40225,
            Duesseldorf, FRG
FEATURES             Location/Qualifiers
     source          1..119704
                     /organism="Odontella sinensis"
                     /organelle="plastid:chloroplast"
                     /db_xref="taxon:2839"
   gene              69086..70546
                     /gene="ycf24"
     CDS             69086..70546
                     /gene="ycf24"
                     /codon_start=1
                     /transl_table=11
                     /product="ORF486"
                     /protein_id="CAA91687.1"
                     /db_xref="GI:1185204"
                     /db_xref="SWISS-PROT:P49530"
                     /translation="MTNKSNKILNTNITKLVNQPYKYGFSTVIEKDIIEKGLNEDVIC
                     LISKKKNEPKFLLEFRLKAFKKWKEMKCPDWAQIKFSEIDYQDIIYYSAPKVKKKLNS
                     LDEVDPELLKTFEKLGISLTEQKRLANVAIDAVFDSVSIATTFKEELAECGVIFSSIS
                     EAIQEYPELIEKYLGSVVPIGDNYFSALNSAVFTDGSFCYIPKDTICPLELSTYFRIN
                     DQKSGQFERTLIVAEKNSQVSYLEGCTAPQYDSNQLHAAVVELVALENADIKYSTVQN
                     WYAGNNYGEGGIYNFVTKRGLCAGSNSKISWTQVETGSNITWKYPSCLLVGDKAKGEF
                     YSVALTNNYQQADTGSKMIHVGKNTRSRIVSKGISAGNSKNTYRGLVNISNKAIGARN
                     YSQCDSLLIGNLSNANTFPFISVQNPTAKIEHEASTSKIGEEQIFYFLQRGIPIEKGV
                     ELMISGFCQEVFTELPLEFAAEADRLLTLKLEGSVG"
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
17:: AJ132267. Skeletonema costa...[gi:4210403]

Links


LOCUS       SCO132267               542 bp    DNA     linear   PLN 29-MAR-2001
DEFINITION  Skeletonema costatum chromoplast ycf24 gene, partial.
ACCESSION   AJ132267
VERSION     AJ132267.1  GI:4210403
KEYWORDS    ycf24 gene.
SOURCE      chromoplast Skeletonema costatum
  ORGANISM  Skeletonema costatum
            Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
            Thalassiosirophycidae; Thalassiosirales; Skeletonemataceae;
            Skeletonema.
REFERENCE   1
  AUTHORS   Tada,N., Otsuka,S., Oyaizu,H. and Matsumoto,S.
  TITLE     Plastid DNA sequences of Skeletonema costatum NIES 323
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 542)
  AUTHORS   Otsuka,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-1999) Otsuka S., Applied Biological Chemistry,
            The University of Tokyo, Yayoi, Bunkyo-ku, Tokyo 113-8657, JAPAN
FEATURES             Location/Qualifiers
     source          1..542
                     /organism="Skeletonema costatum"
                     /organelle="plastid:chromoplast"
                     /strain="NIES 323"
                     /db_xref="taxon:2843"

...
  gene            1..542
                     /gene="ycf24"
     CDS             <1..>542
                     /gene="ycf24"
                     /codon_start=2
                     /transl_table=11
                     /product="Ycf24 protein"
                     /protein_id="CAA10636.1"
                     /db_xref="GI:4210404"
                     /db_xref="SPTREMBL:O96815"
                     /translation="AVFTDGSFCYIPKDVICPLDLSTYFRINDQNSGQFERTLIIAEE
                     NSKVSYLEGCTAPQYDNNQLHAAIVELIALKNATIKYSTVQNWYSGDQKGQGGVYNFV
                     TKRGLCAGDFSKISWTQVETGSSITWKYPSCVLVGDSAQGEFYSVALTNNYQQADTGT
                     KMIHIGRNTRSRIVSKGISA"
BASE COUNT      186 a     74 c    109 g    173 t
ORIGIN
        1 tgcagttttt acagatggat ctttttgtta tataccaaaa gatgttattt gtcctctaga
       61 tttatcaact tattttagga taaatgatca aaactctggt caatttgagc ggacattgat
      121 tatagcagaa gaaaatagta aagtgagtta tttagaagga tgtacagctc ctcaatatga
      181 taataatcaa ctgcatgccg caattgtaga attaattgct ttaaaaaatg caacaataaa
      241 atattctaca gttcaaaatt ggtattctgg ggaccaaaaa gggcaaggtg gtgtgtataa
      301 ctttgtaaca aaacgaggtc tgtgtgcagg tgatttttct aagatttcgt ggactcaagt
      361 tgaaactggt tcttcaataa cgtggaaata tcctagttgt gtactagtgg gtgatagtgc
      421 acaaggagaa ttttattcag ttgctttaac aaataactat caacaagctg acactggtac
      481 aaaaatgatt catataggaa gaaatactcg aagccgtata gtttctaaag ggatttctgc
      541 ag
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

18: AF022186. Cyanidium caldari...[gi:6466296]

Links

```
LOCUS       AF022186              164921 bp    DNA     circular PLN 14-DEC-2000
DEFINITION  Cyanidium caldarium strain RK1 chloroplast, complete genome.
ACCESSION   AF022186 Z36235 Z70297
VERSION     AF022186.2  GI:6466296
KEYWORDS    .
SOURCE      chloroplast Cyanidium caldarium
  ORGANISM  Cyanidium caldarium
            Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales;
            Porphyridiaceae; Cyanidium.
REFERENCE   1  (bases 130696 to 132364)
  AUTHORS   Vogel,H., Fischer,S. and Valentin,K.
  TITLE     A model for the evolution of the plastid sec apparatus inferred
            from secY gene phylogeny
  JOURNAL   Plant Mol. Biol. 32 (4), 685-692 (1996)
  MEDLINE   97134960
   PUBMED   8980520
REFERENCE   2  (bases 1 to 164921)
  AUTHORS   Glockner,G., Rosenthal,A. and Valentin,K.
  TITLE     The structure and gene repertoire of an ancient red algal plastid
            genome
  JOURNAL   J. Mol. Evol. 51 (4), 382-390 (2000)
  MEDLINE   20496959
   PUBMED   11040290
REFERENCE   3  (bases 46857 to 47851)
  AUTHORS   Valentin,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-1996) Institute for Plant Physiology, Justus
            Liebig University, Heinrich Buff Ring 58-62, Giessen 35392, Germany
REFERENCE   4  (bases 28701 to 75580)
  AUTHORS   Gloeckner,G., Rosenthal,A. and Valentin,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Department of Genome Analysis, IMB Jena,
            Beutenbergstr.11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 164921)
  AUTHORS   Gloeckner,G., Rosenthal,A. and Valentin,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1999) Genome Analysis, Institute for Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE   6  (bases 130696 to 132364)
  AUTHORS   Vogel,H., Fischer,S. and Valentin,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-NOV-1999) Institute for Plant Physiology, Justus
            Liebig University, Heinrich Buff Ring 58-62, Giessen 35392, Germany
COMMENT     On or before Nov 23, 1999 this sequence version replaced gi:529651,
            gi:1240002, gi:2465730.
FEATURES             Location/Qualifiers
     source          1..164921
                     /organism="Cyanidium caldarium"
                     /organelle="plastid:chloroplast"
                     /strain="RK1"
                     /db_xref="taxon:2771"

gene            114386..115837
                     /gene="ycf24"
```

APPENDIX D

Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
                /note="similar to ycf24 in Porphyra purpurea"
mRNA            114386..115837
                /gene="ycf24"
                /product="unknown"
CDS             114386..115837
                /gene="ycf24"
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAF12948.1"
                /db_xref="GI:6466366"
                /translation="MIDRKKSSNIQNILNKPYKYGFSTEIQSEEFPKGINEEIIRLMS
                HKKQEPDFILKFRLKAYQIWKKMQAPDWGHLHHNEINFNDVLCYASPKLEQGKNKAQT
                ISEEILATFEKLGVPIKPNNKQPKIAVDAVFDSISFGTTLQKELKEQGIIFCSISEAI
                KAYPNLIKKYLGSIVPAGDNYFAALNSAVFTDGSFCYIPKNIRCPVDLSTYFRINNKE
                AGQFERTLIIADENSFVNYLEGCTAPQFDTNQLHAAVVELICFKNATINYSTVQNWYA
                GNNKGEGGVYNFVTKRGLCQGENSKISWTQLETGSAITWKYPSCLLKGKRSTGEFFSV
                TLTNNAQEADTGTKMLHFGRQSKSLVISKGISGGVSKNTYRGLVKISGSAIYSDNRSQ
                CDSLLIGKGSESNTYPNLHVHNSLSKVEHEAFVSRIGEEQIFYFQQRGINIEEALNMI
                VSGFCQDVCNKLPMEFALEANKLLNIKLEGSIG"
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

19: AF138960. Neospora caninum ...[gi:6492292]

Links

```
LOCUS       AF138960               3343 bp    DNA     linear   INV 01-DEC-1999
DEFINITION  Neospora caninum ycf24 protein (ycf24) gene, partial cds; DNA
            dependent RNA polymerase beta subunit (rpoB) gene, complete cds;
            and DNA dependent RNA polymerase beta subunit' (rpoC1) gene,
            partial cds, plastid genes for plastid products.
ACCESSION   AF138960
VERSION     AF138960.1  GI:6492292
KEYWORDS    .
SOURCE      plastid Neospora caninum
  ORGANISM  Neospora caninum
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Sarcocystidae; Neospora.
REFERENCE   1  (bases 1 to 3343)
  AUTHORS   Lang-Unnasch,N. and Aiello,D.P.
  TITLE     Sequence evidence for an altered genetic code in the Neospora
            caninum plastid
  JOURNAL   Int. J. Parasitol. 29 (10), 1557-1562 (1999)
  MEDLINE   20074141
   PUBMED   10608442
REFERENCE   2  (bases 1 to 3343)
  AUTHORS   Lang-Unnasch,N. and Aiello,D.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-1999) Div. Geographic Medicine, Dept. Medicine,
            University of Alabama at Birmingham, 845 S. 19th Street,
            Birmingham, AL 35294-2170, USA
FEATURES             Location/Qualifiers
     source          1..3343
                     /organism="Neospora caninum"
                     /organelle="plastid"
                     /strain="Nc1"
                     /db_xref="taxon:29176"
                     /note="the apicoplast is a vestigial nonphotosynthetic
                     plastid
                     plastid: apicoplast"
     gene            <1..43
                     /gene="ycf24"
     CDS             <1..43
                     /gene="ycf24"
                     /note="Toxoplasma gondii Ycf24 homolog"
                     /codon_start=2
                     /product="ycf24 protein"
                     /protein_id="AAF14260.1 [appellants note: pllivaraqk lfy]"
                     /db_xref="GI:6492293"
                     /translation="PLLIVARAQKLFY"
     gene            75..3227
                     /gene="rpoB"
     CDS             75..3227
                     /gene="rpoB"
                     /note="Toxplasma gondii RpoB and Plasmodium falciparum
                     RpoB homolog"
                     /codon_start=1
                     /transl_except=(pos:222..224,aa:Trp)
                     /transl_except=(pos:1866..1868,aa:Trp)
                     /transl_except=(pos:2976..2978,aa:Trp)
```

APPENDIX D

Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
                     /product="DNA dependent RNA polymerase beta subunit"
                     /protein_id="AAF14261.1"
                     /db_xref="GI:6492294"
                    /translation="MAFKIKTFLSIKSHKIFYTDFYFFLKKKLIILIKNIFPEYFNFN
                     NKYKNWNLICLPDFLYFKVNNTHFLDNVQYINSLIKIFLPLKFQNLKTNQIFFKNLLI
                     FELPKYNSHNYCYLNGLKKIFISKYFTSNGIFFHKYLKRKYDIYYAKILLTNSNFFNI
                     VLDLKLKQIYLSIKNLKFNFILFLYYLGINNKDILKYSRYKKSKILKLLIFTALQTNL
                     INNKKFILKNLNYLKSIFKVTNLNKNYKHFIINKNKLNYNYGNFNKNNFLTIDFIFIL
                     DLLLDLQSKKLCFKNIDHLDNKHINTIGNYFQHNFKFYLKKFISIIPNLIKLKKFSLL
                     KVYNFKELLILNPLIQYLEQINSFSELTHKYKLNNYNSSLKGILNLREICLNQIGKLC
                     LIDTTEGINCGLIVNFAKHIRIYKKGIIQIYVSPVFKNKTKEFINFKTSLDQELYLIQ
                     FNNINLRKNKLFNENIKVVYNKNNFRIKFISNKKSFLLKFADLFSFTENLIPFIKYND
                     PARCLMGAKMQSQSVPLLNKKKSFVVTGYEKEIITKSDITIKALQEGIVLNASSLKIH
                     IKDLFNREIVYYLSKYKKSNQNTLIHQKPLVWNGERVFTNQLLTQHQDIIDSEFAIGN
                     NLLIYYGNFCGYDFEDAVIGSKRVLYQQLFSSLHMDIYEFNFGYNNENDIEFSTLEIP
                     KQSYYIKKSLDSLGIVKEGEKILTGNILLTKIKVTKPNYTYKSIFKLIYSIFGKTIRN
                     IKDNSLYIQTGKNGRVSKIELFLINTNSHYKTYNNSYLKCRIFICKQRFLTVGDKLCG
                     RYGNKGILSYIAENADLPFLQNSFYPDIIVGALGIPSRMNLGQLFEALVGKIGFSYNI
                     RILPSFTSSSNAYFNYLKILIYNFLMFNNLKKGFNWLYNFNLPGKFLIRDGRTGIKLK
                     SSVLCGVSRYSKLIHMIKDKLHFRTTGPYTEILQQPLKGKKNLGGQRFGEMEIWALEA
                     FGASYNLKEILNYKSDDCFARNNLKDYLLFRNSELQNSTITESFRVILKEFNGLILNL
                     ELFLITDDLEENYLNLTINY"
     gene            3243..>3343
                     /gene="rpoC1"
     CDS             3243..>3343
                     /gene="rpoC1"
                     /note="Toxoplasma gondii RpoC1 homolog"
                     /codon_start=1
                     /transl_except=(pos:3318..3320,aa:Trp)
                     /product="DNA dependent RNA polymerase beta subunit'"
                     /protein_id="AAF14262.1"
                     /db_xref="GI:6492295"
                     /translation="MKKIYIKNNTIGFRLSLASPNLIIKWSLKYIKN"
BASE COUNT     1416 a    272 c    306 g   1349 t
ORIGIN
        1 tcctctttta attgtagcta gagcacaaaa attattttat taatttttg aatttaatgt
       61 aatttattaa aattatggct tttaaaataa aaacttttt gtcaattaaa tcacataaaa
      121 ttttttatac tgatttttat ttttttttaa aaaaaaaatt aattatttta ataaaaaata
      181 tttttccaga atattttaat tttaataata aatataaaaa ttgaaattta atttgccttc
      241 ctgacttctt atattttaaa gttaataata ctcatttttt agataatgtt caatatataa
      301 attctttaat aaaaatattt ttaccattaa aatttcaaaa tttaaaaaca aatcaaattt
      361 tttttaaaaa tttattaatt tttgagttac caaaatataa ttcacataat tattgttatt
      421 taaatggttt aaaaaaaatt tttatttcaa aatattttac ttcaaatggt atctttttc
      481 ataaatattt aaaaagaaaa tatgatattt attatgctaa aatattatta actaattcaa
      541 attttttaa tattgtttta gatttaaaat tgaaacaaat ctatttatct attaaaaatt
      601 taaaatttaa ttttatttta tttttatatt atttaggaat taataataaa gatattttaa
      661 aatattctag atataaaaaa tctaaaattt taaaattatt aatttttaca gctttacaaa
      721 ctaatttaat taataataaa aaatttatat taaaaaatct aaattattta aaatctattt
      781 ttaaagtaac aaatttaaat aaaaattata aacattttat aattaataaa aacaaattaa
      841 attataatta tggaaatttt aataaaaata attttttaac tattgatttt atatttatat
      901 tagatttatt attagattta caatctaaaa aattatgttt taaaaatata gatcatttag
      961 ataataaaca tattaataca ataggtaatt attttcaaca taattttaaa ttctatttaa
     1021 aaaaatttat atctattata ccaaatttaa ttaaattaaa aaaatttttct ttactaaaag
     1081 tatataattt taaagaatta ttaattctta acccattaat tcaatattta gaacaaatta
     1141 atagtttctc tgaacttacc cataaatata aattaaataa ttataattct tctttaaaag
     1201 gaattttaaa tctacgagaa atttgtttaa atcaaatagg taaactatgt ttaatagata
     1261 ctacagaagg aattaattgt ggtttaatag ttaactttgc taaacatatt agaatttata
     1321 aaaaaggcat aatacaaatt tatgttagtc ctgtctttaa aaataaaact aaagaattta
     1381 taaatttaa aacatctcta gatcaagaat tatatttaat tcaatttaat aacattaatt
```

APPENDIX D

Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
1441 tacgtaaaaa taaattattt aatgaaaata ttaaagtagt atataataag aataatttta
1501 gaattaaatt tatttccaat aaaaaaagtt ttttattaaa atttgcagat ttattttcat
1561 ttacagaaaa tttaatacct tttattaaat ataatgaccc agcaagatgt ttaatgggag
1621 ctaagatgca aagtcaatca gttccattat taaataaaaa aaaatctttt gttgttactg
1681 ggtatgaaaa agaaataata actaaatctg atattactat taaagcttta caagaaggaa
1741 tagttttaaa tgcttcttct ttaaaaatac atattaaaga cctatttaat agagaaatag
1801 tttattattt atcaaaatat aaaaaaagta atcaaaacac attaattcat caaaaaccat
1861 tagtatgaaa tggtgaaaga gtatttacta atcaattatt aacacaacat caagatatta
1921 ttgactcaga atttgcaata ggaaataatt tattgattta ttatggtaat ttttgtggat
1981 atgattttga agatgctgta attggtagta aaagagtttt atatcaacaa ttatttagtt
2041 ctcttcatat ggacatttat gaatttaatt ttgggtataa taatgaaaat gatattgaat
2101 ttagtacatt agaaattcct aaacaaagtt attatattaa aaaaagctta gattctttag
2161 gaattgttaa agaaggtgaa aaaattttaa ctggtaatat tttattaaca aaaataaaag
2221 ttacaaaacc taattataca tataaatcta tatttaaact tatatactct atttttggta
2281 aaacgattag aaatattaaa gataattctt tatatattca aacaggaaaa aatggtagag
2341 taagtaaaat tgaattattt ctaataaata caaattctca ttataaaact tataataaca
2401 gttatttaaa atgtagaatt tttatatgta aacaaagatt tttaactgtt ggagataaat
2461 tatgtggaag atatggaaat aaagggattt tatcgtatat agcagaaaat gctgatttac
2521 ctttttttaca aaattctttt tatccagata ttattgttgg tgctttaggt attccatcac
2581 gtatgaattt aggacaatta tttgaagctt tagttgggaa aataggtttt tcttataata
2641 ttagaatttt accttcattc acttcttctt ctaacgcgta ttttaattat ttaaaaattt
2701 taatatataa ttttttaatg ttcaataatc ttaaaaaagg ttttaattgg ctttataact
2761 ttaatttgcc aggaaaattt cttattagag atggaagaac tggtataaaa ttaaaatctt
2821 cagttctttg tggagtttct agatattcaa aattaattca tatgattaag gataaacttc
2881 attttagaac aacaggacct tatactgaaa ttttacaaca acctttaaaa ggtaaaaaaa
2941 atttaggtgg tcaacgattt ggagaaatgg aaatttgagc tctagaagct tttggtgctt
3001 catataattt aaaagaaatt ttaaattata aatctgatga ttgtttcgca cgtaataatc
3061 tgaaagatta tttattattt agaaatagtg aattacaaaa ttcaactata actgaatctt
3121 ttcgtgttat tttaaaagaa tttaatggat taattttaaa tttagaatta tttttaataa
3181 cggacgattt agaagaaaat tatcttaatt taactattaa ttattaataa ttaaaattta
3241 tcatgaaaaa aatttatata aaaaacaata caataggttt tcgactatca ttagcttctc
3301 caaatttaat aattaaatga tctttaaaat atataaaaaa ttt
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

20: U87145. Toxoplasma gondii...[gi:5231237]

Links

```
LOCUS       U87145                 34996 bp    DNA     circular INV 29-JUN-1999
DEFINITION  Toxoplasma gondii chloroplast, complete genome.
ACCESSION   U87145
VERSION     U87145.2  GI:5231237
KEYWORDS    .
SOURCE      chloroplast Toxoplasma gondii
  ORGANISM  Toxoplasma gondii
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Sarcocystidae; Toxoplasma.
REFERENCE   1  (bases 4337 to 4925)
  AUTHORS   Beckers,C.J., Roos,D.S., Donald,R.G., Luft,B.J., Schwab,J.C.,
            Cao,Y. and Joiner,K.A.
  TITLE     Inhibition of cytoplasmic and organellar protein synthesis in
            Toxoplasma gondii. Implications for the target of macrolide
            antibiotics
  JOURNAL   Journal of Clin. Invest. 95, 367-376 (1995)
REFERENCE   2  (bases 13791 to 14996)
  AUTHORS   Kohler,S., Delwiche,C.F., Denny,P.W., Tilney,L.G., Webster,P.,
            Wilson,R.J., Palmer,J.D. and Roos,D.S.
  TITLE     A plastid of probable green algal origin in Apicomplexan parasites
  JOURNAL   Science 275 (5305), 1485-1489 (1997)
  MEDLINE   97197911
   PUBMED   9045615
REFERENCE   3  (bases 13218 to 15870; 28669 to 34697)
  AUTHORS   Denny,P., Preiser,P., Williamson,D. and Wilson,I.
  TITLE     Evidence for a single origin of the 35 kb plastid DNA in
            Apicomplexans
  JOURNAL   Protist 149, 51-59 (1998)
REFERENCE   4  (bases 1 to 34996)
  AUTHORS   Kissinger,J.C., Donald,R.G., Moulton,A.L., Gutell,R., Aiello,D.P.,
            Lang-Unnasch,N. and Roos,D.S.
  TITLE     Mapping, cloning, and complete sequence annotation of the 35-kb
            plastid genome of Toxoplasma gondii
  JOURNAL   Unpublished
REFERENCE   5  (bases 1 to 34996)
  AUTHORS   Kissinger,J.C., Donald,R.G., Moulton,A.L., Aiello,D.P.,
            Lang-Unnasch,N. and Roos,D.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Biology, University of Pennsylvania, 415 S.
            University Ave, Philadelphia, PA 19104, USA
REFERENCE   6  (bases 1 to 34996)
  AUTHORS   Kissinger,J.C., Donald,R.G., Moulton,A.L., Aiello,D.P.,
            Lang-Unnasch,N. and Roos,D.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1999) Biology, University of Pennsylvania, 415 S.
            University Ave, Philadelphia, PA 19104, USA
  REMARK    Sequence update by submitter
COMMENT     On Jun 28, 1999 this sequence version replaced gi:1870709.
            This updated entry provides annotation (and corrected sequence
            information) for the entire 35kb apicoplast genome of Toxoplasma
            gondii, based on extensive resequencing from independent molecular
            clones and a comprehensive reanalysis of the entire sequence.
            Unlike the apicoplast genome of Plasmodium falciparum (Genbank
            accession numbers X95275 & X95276) the T. gondii apicoplast genome
```

APPENDIX D

Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
            exhbits non-standard codon usage: 33 in-frame UGA codons interrupt
            17 of the 28 predicted coding regions, and are presumed to encode
            tryptophan.  In addition, two genes (rps8 and rpoC2) contain
            in-frame UAA or UAG stop codons which are currently being
            investigated further.  Note that the identification of many genes
            is hypothetical, based in part on the known organization of
            ribosomal protein superoperons in other taxa.
FEATURES             Location/Qualifiers
     source          1..34996
                     /organism="Toxoplasma gondii"
                     /organelle="plastid:chloroplast"
                     /db_xref="taxon:5811"
..

     CDS             complement(28289..29686)
                     /note="in frame UGA codon predicted to encode tryptophan;
                     similar to Plasmodium falciparum plastid genome ORF470"
                     /codon_start=1
                     /transl_except=(pos:complement(28916..28918),aa:Trp)
                     /product="ycf24 homolog"
                     /protein_id="AAD41153.1"
                     /db_xref="GI:5231259"
                     /translation="MKLYKYLYNKYNNNTDLFNTVRLIGGLNINMVNKLIFKQDNFIF
                     LYIFRLNALSILNKFKQPDWCFYELPEFAFDDISYYSIPLNVYTNKNKYKSILSKLGL
                     ELKFSENLILDVIFDSVLLLNLTTFFLIKMGLFFLSFFQSIIFYPYLIFSYLGSIVSN
                     TDNFFLTINSIIFNEGSFCFVMKDLNSNINLTTYFRTHSENFAQFERTLIVLSENSKL
                     IYFEGCSAPMFLESQLHIAIVELFIKTKANLKYSTIQNWYRGNQLGEGGLYNFTTKRG
                     FCMDKSFLNWIQIEIGSVITWKYPSTYLIGNKSFSNFFSLAMLSDYQVSDTGTKMLHI
                     GKNTKSFILSKSLSFNFSFYTYRGLVTIFKTALNSYNYTECNSLLIGCNAFTATIPYT
                     IINNFSAYINQEATISKLELDFLFFLLHRGLNLKSTLMILIYGYCYNICSKISFELEL
                     EVPLLIVARAQKLFY"
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
21:AF095904. Toxoplasma gondii...[gi:4336507]

Links


LOCUS       AF095904                3360 bp    DNA     linear   INV 04-MAR-1999
DEFINITION  Toxoplasma gondii ycf24 protein (ycf24) gene, partial cds; DNA
            dependent RNA polymerase beta subunit (rpoB) gene, complete cds;
            and DNA dependent RNA polymerase beta' subunit (rpoC1) gene,
            plastid genes encoding plastid proteins, partial cds.
ACCESSION   AF095904
VERSION     AF095904.1  GI:4336507
KEYWORDS    .
SOURCE      plastid Toxoplasma gondii
  ORGANISM  Toxoplasma gondii
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Sarcocystidae; Toxoplasma.
REFERENCE   1  (bases 1 to 3360)
  AUTHORS   Aiello,D.P. and Lang-Unnasch,N.
  TITLE     Analysis of the rpoB gene product of Toxoplasma gondii
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3360)
  AUTHORS   Aiello,D.P. and Lang-Unnasch,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-1998) Geographic Medicine, University of Alabama
            at Birmingham, 845 South 19th Street, Birmingham, AL 35294-2170,
            USA
FEATURES             Location/Qualifiers
     source          1..3360
                     /organism="Toxoplasma gondii"
                     /organelle="plastid"
                     /strain="RH"
                     /db_xref="taxon:5811"
                     /note="the apicoplast is a vestigial nonphotosynthetic
                     plastid
                     plastid: apicoplast"
     gene            <1..43
                     /gene="ycf24"
     CDS             <1..43
                     /gene="ycf24"
                     /note="ORF 470; similar to ORF 470 of Plasmodium
                     falciparum apicoplast and Porphyra pupurea plastid"
                     /codon_start=2
                     /product="ycf24 protein"
                     /protein_id="AAD17841.1  [appellants note: pllivaraqk lfy]"
                     /db_xref="GI:4336508"
                     /translation="PLLIVARAQKLFY"
     gene            74..3229
                     /gene="rpoB"
     CDS             74..3229
                     /gene="rpoB"
                     /EC_number="2.7.7.6"
                     /codon_start=1
                     /transl_except=(pos:221..223,aa:Trp)
                     /transl_except=(pos:1862..1864,aa:Trp)
                     /transl_except=(pos:2750..2752,aa:Trp)
                     /product="DNA dependent RNA polymerase beta subunit"
                     /protein_id="AAD17842.1"
                     /db_xref="GI:4336509"
```

APPENDIX D

Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
                     /translation="MNFKIKTFFSIKSNKIFYTDFYFFLKKKLIILIKNIFPEYFNFN
                     NKYKNWNLICLYDLLYFKVNNINFLDNVQYINSLIKIFLPLKFQNLKTNKVFFKNLLI
                     FELPKYNSYNYCYLNGLKKIFISKYFTSNGIFFNKYLKRKYNIYYAKLLLTNSNFFNI
                     ILDLKLKQIYLSIKNLKFNFILFLYYLGIKNTDILKYSRYKNSKILKLLIFSALQTDL
                     VNNKKIILKNLNYLKSIFKLTILNKNYKNFIISKDKLNYNYGDFSKNNLLIIDFIFIL
                     DLLLDLQSKKLCFKTIDHLDNKHINTIGNYFQHNFKFYLKKFISIIPNLIKLQKFSLL
                     KVYNFKELLILNPLIQYLEQINSFSELMHKYKLNNYNSFSKGILNLREICLNQLGKLC
                     LIDTTEGINCGLVVSFAKHIRIYKKGIIQVYFSSIFKNKTKEFLNFKTSLDQELYLIQ
                     FNNINLRKIKYLMNVRLVYNKNNFRIKFFSNKKSILLEFTDLFSFTENLIPFIKYNDP
                     ARCLMGAKMQSQSVPLLNKKKSFVITGYEKEIITKSDTTIKALQEGIVLNASSLKIHI
                     KDLFNREIVYYLSKYKKSNQNTIIHQKPLVWNGERVFTNQLLTQHQDIIDSEFAIGNN
                     LLIYYGNFCGYDFEDAVIVSKRVLYQQLFSSLHMDIYEFNFCYNNENDIEFSTLEIPK
                     QSYYIKKNLDSLGIIKEGEKILTGSILLTKIKVAKPTYTYKSIFKLIYSIFGKTIRNI
                     KDNSLYIQTGKSGRVSKIELFLVNISSRHKYKTYNNSYLKCRIFICKQRFLTVGDKLC
                     GRYGNKGILSYIAENADLPFLQNSFYPDIIVGALGIPSRMNLGQLFEALVGKISFSYN
                     IRILPSFTTSSNLYFNYLKILIYNFLMFNNFKKGFNWLYNFNLPGKFIIRDGRTGVKL
                     KSSVLCGVSRYSKLIHLIKDKLHFRTTGPYTEILQQPLKGKKNLGGQRFGEMEIWALE
                     AFGASYNLKEILNYKSDDCFARNNLKEYLLFRNTELQNSTITESFRVILKEFNGLILN
                     LELFLITDDLEENYLNLTINY"
     gene            3249..>3360
                     /gene="rpoC1"
     CDS             3249..>3360
                     /gene="rpoC1"
                     /codon_start=1
                     /transl_except=(pos:3324..3326,aa:Trp)
                     /product="DNA dependent RNA polymerase beta' subunit"
                     /protein_id="AAD17843.1"
                     /db_xref="GI:4336510"
                     /translation="MKKIFIKNNTIGFRLSLASPNLIIKWSLKYIKKFFYF"
BASE COUNT     1427 a    254 c    299 g   1380 t
ORIGIN
        1 tcctctttta attgtagcaa gagcacaaaa attattttat taatttttta ttttaataaa
       61 atttttaaa attatgaatt ttaaaataaa gacttttttt tcaattaaat caaataaaat
      121 tttttatact gatttttatt tttttttaaa aaaaaaatta attattttaa taaaaaatat
      181 atttccagaa tattttaatt ttaataataa atataaaaat tgaaatttaa tttgtcttta
      241 tgacttatta tattttaaag ttaataatat taatttttta gataatgtac aatatataaa
      301 ctctttaata aaaatttttt taccattaaa atttcaaaat ttaaaaacaa ataaagtttt
      361 ttttaagaat ttattaattt ttgaattacc taaatataat tcatataatt attgttattt
      421 aaatggttta aaaaaaatat ttatttcaaa atattttact tcaaatggta tatttttaa
      481 taaatattta aaaagaaaat ataatattta ttatgctaaa ttattattaa caaattcaaa
      541 tttttttaat attattttag atttaaaatt aaaacaaatt tatttatcta ttaaaaattt
      601 aaaatttaat tttatttttat ttttatatta tttaggaatt aaaaatacag atattttaaa
      661 atattctaga tataaaaatt ctaaaatttt aaaattatta atttttcag ctttacaaac
      721 tgatttagtt aataataaaa aaattatatt aaaaaattta aattatttaa aatctatttt
      781 taaattaaca attttaaata aaaattataa aaatttttata attagtaaag ataaattaaa
      841 ttataattac ggagatttta gtaaaaataa tcttttaatt attgatttta tatttatatt
      901 agatttatta ttagatttac aatctaaaaa attatgtttt aaaactatag atcatttaga
      961 taataaacac attaatacaa taggtaatta ttttcaacat aattttaaat tttatttaaa
     1021 aaaatttata tctattatac caaatttaat caaattacaa aaatttttctt tattaaaagt
     1081 atataatttt aaagaattat taattcttaa cccattaatt caatatttag aacaaattaa
     1141 tagtttctca gaacttatgc ataaatataa attaaataat tataattctt tttcaaaagg
     1201 aattttaaat ttacgagaaa tttgtttaaa tcaattaggt aaattatgtt taatagatac
     1261 taccgaagga attaattgtg gtttagttgt tagttttgct aaacatatac gaatatataa
     1321 aaaaggtata atacaagttt attttagttc tatatttaaa aataaaacta aagaattttt
     1381 aaatttttaaa acatctttag atcaagaatt atatttaatt caatttaata atattaattt
     1441 acgtaaaata aaatatttaa tgaacgttag attagtatat aataaaaata attttagaat
     1501 taaatttttt tctaataaaa aaagtatttt attagaattt acagatttat tttcatttac
     1561 agaaaattta attcctttta ttaaatataa cgacccagca agatgcttaa tgggagctaa
     1621 aatgcaaagt caatcagttc cattattaaa taaaaaaaaa tcttttgtta ttactggata
```

APPENDIX D

Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
1681 tgaaaaagaa ataataacta aatcagatac tactattaaa gctctacaag aaggaatagt
1741 tttaaatgct tcttctttaa aaatacatat taaagattta tttaatagag agatagtata
1801 ttatttatca aaatataaaa aaagtaatca aaatacaata attcatcaaa aacctttagt
1861 atgaaacggt gaaagagtat ttactaatca attactaaca caacatcaag atattataga
1921 ctcagaattt gcaataggaa ataatttatt aatttattat ggtaattttt gtggatatga
1981 ttttgaagat gctgtaattg ttagtaaaag agttttatat caacaattat tcagttctct
2041 tcatatggat atttatgaat ttaatttttg ttataataat gaaaatgaca ttgaatttag
2101 tacattagaa atccctaaac aaagttatta tattaaaaaa aatttagatt ccttaggtat
2161 tattaaagaa ggtgaaaaaa ttttaactgg aagtatttta ctaacaaaaa taaaagttgc
2221 aaaacctact tatacttata aatctatatt taaacttata tattctattt ttggtaaaac
2281 aattagaaat attaaagata attctttata tattcaaaca ggaaaaagcg gtagagtaag
2341 taaaattgaa ttatttttag taaatataag ttctcgtcat aaatataaaa cttataataa
2401 tagttattta aaatgtagaa tttttatatg taaacaaaga tttttaacag ttggagataa
2461 attatgtgga agatatggaa ataaaggaat tttatcttat atagcagaaa acgctgattt
2521 acctttttta caaaattctt tttatccaga tattattgtt ggtgccttag gtattccatc
2581 tcgcatgaat ttaggacaat tatttgaagc tttagttggg aaaataagtt tttcttataa
2641 tattagaatt ttaccttcat ttactacttc ttctaattta tattttaatt atctaaaaat
2701 tttaatatat aatttttaa tgtttaataa ctttaaaaaa ggttttaatt gattatataa
2761 ttttaattta ccaggaaaat ttattattag agatggaaga actggtgtaa aattaaaatc
2821 ttcggttctt tgtggagttt ctaggtattc aaaattaatt catttgatta aagataaact
2881 tcattttaga actacaggac cttatactga aattttacaa caacctttaa aaggtaaaaa
2941 aaatttagga ggtcaacgat ttggagaaat ggagatttgg gctttagaag cttttggagc
3001 ttcatataat ttaaaagaaa ttttaaatta taaatctgat gattgttttg cacgtaataa
3061 tcttaaagaa tatttattat ttagaaatac tgaactacaa aattcaacta ttactgaatc
3121 ttttcgcgtt attttaaaag aatttaatgg attaatttta aatttagaat tatttttaat
3181 aacagacgat ttagaagaaa attatcttaa tttaactatt aactattaat aaataattaa
3241 aatttattat gaaaaaaatc tttataaaaa ataatacaat aggttttaga ttatcattag
3301 cttctcctaa tttaataatt aaatgatctt taaaatatat aaaaaaattt ttttatttta
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

22: AF041468. Guillardia theta ...[gi:3602932]

Links

```
LOCUS       AF041468              121524 bp    DNA     circular PLN 03-MAR-1999
DEFINITION  Guillardia theta complete plastid genome.
ACCESSION   AF041468 X14171 X62349 X51511 X14504 X52158 X52912 X56806 M76547
            X62348 Z21976 U81044 AF063017
VERSION     AF041468.1  GI:3602932
KEYWORDS    .
SOURCE      chloroplast Guillardia theta
  ORGANISM  Guillardia theta
            Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
REFERENCE   1  (bases 47701 to 48415)
  AUTHORS   Douglas,S.E. and Durnford,D.G.
  TITLE     The small subunit of ribulose-1,5-bisphosphate carboxylase is
            plastid-encoded in the chlorophyll c-containing alga Cryptomonas
            phi
  JOURNAL   Plant Mol. Biol. 13 (1), 13-20 (1989)
  MEDLINE   93357429
   PUBMED   2562756
REFERENCE   2  (bases 18535 to 19351)
  AUTHORS   Douglas,S.E. and Durnford,D.G.
  TITLE     Sequence analysis of the plastid rDNA spacer region of the
            chlorophyll c-containing alga Cryptomonas phi
  JOURNAL   DNA Seq. 1 (1), 55-62 (1990)
  MEDLINE   92119320
   PUBMED   2132959
REFERENCE   3  (bases 43739 to 44938)
  AUTHORS   Douglas,S.E. and Durnford,D.G.
  TITLE     Nucleotide sequence of the genes for ribosomal protein S4 and
            tRNA(Arg) from the chlorophyll c-containing alga Cryptomonas phi
  JOURNAL   Nucleic Acids Res. 18 (7), 1903 (1990)
  MEDLINE   90245597
   PUBMED   2336372
REFERENCE   4  (bases 34539 to 35380)
  AUTHORS   Reith,M. and Douglas,S.
  TITLE     Localization of beta-phycoerythrin to the thylakoid lumen of
            Cryptomonas phi does not involve a signal peptide
  JOURNAL   Plant Mol. Biol. 15 (4), 585-592 (1990)
  MEDLINE   91338697
   PUBMED   2102376
REFERENCE   5  (bases 45872 to 47981)
  AUTHORS   Douglas,S.E., Durnford,D.G. and Morden,C.W.
  TITLE     Nucleotide sequence of the gene for the large subunit of
            ribulose-1,5-bisphosphate carboxylase/oxygenase from the
            chlorophyll c-containing Alga Cryptomonas F: evidence supporting
            the polyphyletic origin of plastids
  JOURNAL   J. Phycol. 26, 500-508 (1990)
REFERENCE   6  (bases 110917 to 113854)
  AUTHORS   Douglas,S.E.
  TITLE     Unusual organization of a ribosomal protein operon in the plastid
            genome of Cryptomonas phi: evolutionary considerations
  JOURNAL   Curr. Genet. 19 (4), 289-294 (1991)
  MEDLINE   91330343
   PUBMED   1868578
REFERENCE   7  (bases 40675 to 42376)
  AUTHORS   Douglas,S.E. and Turner,S.
```

APPENDIX D

Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
  TITLE     Molecular evidence for the origin of plastids from a
            cyanobacterium-like ancestor
  JOURNAL   J. Mol. Evol. 33 (3), 267-273 (1991)
  MEDLINE   92099311
   PUBMED   1757997
REFERENCE   8  (bases 96129 to 98906)
  AUTHORS   Wang,S.L. and Liu,X.Q.
  TITLE     The plastid genome of Cryptomonas phi encodes an hsp70-like
            protein, a histone-like protein, and an acyl carrier protein
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 88 (23), 10783-10787 (1991)
  MEDLINE   92073372
   PUBMED   1961745
REFERENCE   9  (bases 106789 to 108216)
  AUTHORS   Douglas,S.E.
  TITLE     A secY homologue is found in the plastid genome of Cryptomonas phi
  JOURNAL   FEBS Lett. 298 (1), 93-96 (1992)
  MEDLINE   92183838
   PUBMED   1544427
REFERENCE   10 (bases 42198 to 44153)
  AUTHORS   Douglas,S.E. and Reith,M.E.
  TITLE     A bchI homolog, encoding a subunit of Mg chelatase, is located on
            the plastid genomes of red and cryptomonad algae
  JOURNAL   J. Mar. Biotechnol. 1, 135-141 (1993)
REFERENCE   11 (bases 82327 to 84479)
  AUTHORS   Douglas,S.E. and Murphy,C.A.
  TITLE     Structural, transcriptional and phylogenetic analyses of the atpB
            gene cluster from the plastid of Cryptomonas F (Cryptophyceae)
  JOURNAL   J. Phycol. 30, 329-340 (1994)
REFERENCE   12 (bases 98901 to 114602)
  AUTHORS   Wang,S.L., Liu,X.Q. and Douglas,S.E.
  TITLE     The large ribosomal protein gene cluster of a cryptomonad plastid:
            gene organization, sequence and evolutionary implications
  JOURNAL   Biochem. Mol. Biol. Int. 41 (5), 1035-1044 (1997)
  MEDLINE   97283757
   PUBMED   9137835
REFERENCE   13 (bases 61067 to 68605)
  AUTHORS   Leitsch,C.E.W., Kowallik,K.V. and Douglas,S.E.
  TITLE     The atpA gene cluster of a cryptomonad, Guillardia theta: A piece
            in the puzzle of chloroplast genome development
  JOURNAL   J. Phycol. (1998) In press
REFERENCE   14 (bases 1 to 121524)
  AUTHORS   Douglas,S.E. and Penny,S.L.
  TITLE     The plastid genome of the cryptophyte alga, Guillardia theta:
            complete sequence and conserved synteny groups confirm its common
            ancestry with red algae
  JOURNAL   J. Mol. Evol. 48 (2), 236-244 (1999)
  MEDLINE   99128221
   PUBMED   9929392
REFERENCE   15 (bases 1 to 121524)
  AUTHORS   Douglas,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-1998) Institute for Marine Biosciences, National
            Research Council, 1411 Oxford Street, Halifax, Nova Scotia B3H 3Z1,
            Canada
COMMENT     On or before Sep 15, 1998 this sequence version replaced gi:11396,
            gi:11297, gi:18103, gi:18281, gi:11383, gi:11407, gi:12539,
            gi:336730, gi:11300, gi:398949, gi:2661180.
FEATURES             Location/Qualifiers
     source          1..121524
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
                     /organism="Guillardia theta"
                     /organelle="plastid:chloroplast"
                     /db_xref="taxon:55529"
                     /note="plastid"

...

     gene            61067..62518
                     /gene="ycf24"
      CDS            61067..62518
                     /gene="ycf24"
                     /note="hypothetical chloroplast RF24"
                     /codon_start=1
                     /product="ABC transporter"
                     /protein_id="AAC35664.1"
                     /db_xref="GI:3603003"
                     /translation="MSDDLSKRSLRELVSQPYKYGFHTDIENEEFPKGLDEDIIKEIS
                     KLKCEPSYMLDFRLKSYILWKKMSLPEWACLTYLNINYQDIVYYSAPKNSTKLDSLED
                     VDKKILETFDKLGIPLNEQKKLANVAVDAIFDSVSVGTTFKQELSNVGVLFCPLSEAT
                     NKYSTLVEKYLGSVVPIGDNYFAALNSAVFSEGSFCYIPPNVKCPLELSTYFRINNEN
                     SGQFERTLIIADFNSYVSYLEGCTAPMYDKNQLHAAVVELIALENAEIRYSTVQNWYS
                     GDTNGKGGIYNFVTKRGLCAGKSSKISWTQVETGSAITWKYPSCILVGEDSVGEFYSV
                     ALTNNYQQADTGTKMIHVGRGSKSRIISKGISAGYSKNTYRGQVKININALGSINNSQ
                     CDSMLIGPYSQANTYPYIQVSNAMSRVEHEASTSKIEEEQLFYFLQRGISVEQAISLL
                     ISGFCRDVFVKLPMEFAVEADKLLSVKLEGTVG"
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

23:U38804. Porphyra purpurea...[gi:1276652]

Links

```
LOCUS       PPU38804              191028 bp    DNA     circular PLN 27-MAR-1998
DEFINITION  Porphyra purpurea chloroplast, complete genome.
ACCESSION   U38804
VERSION     U38804.1  GI:1276652
KEYWORDS    .
SOURCE      chloroplast Porphyra purpurea
  ORGANISM  Porphyra purpurea
            Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
            Porphyra.
REFERENCE   1  (bases 1 to 191028)
  AUTHORS   Reith,M.E. and Munholland,J.
  TITLE     Complete nucleotide sequence of the Porphyra purpurea chloroplast
            genome
  JOURNAL   Plant Mol. Biol. Rep. 13 (4), 333-335 (1995)
REFERENCE   2  (bases 1 to 191028)
  AUTHORS   Reith,M.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-1995) Michael E. Reith, Marine Biology Section,
            NRC Institute for Marine Biosciences, 1411 Oxford Street, Halifax,
            Nova Scotia B3H 3Z1, Canada
FEATURES             Location/Qualifiers
     source          1..191028
                     /organism="Porphyra purpurea"
                     /organelle="plastid:chloroplast"
                     /strain="Avonport"
                     /db_xref="taxon:2787"

...
   gene              40948..42411
                     /gene="ycf24"
     CDS             40948..42411
                     /gene="ycf24"
                     /note="hypothetical chloroplast ORF 24."
                     /codon_start=1
                     /protein_id="AAC08126.1"
                     /db_xref="GI:1276706"
                     /translation="MVNTQNQISQTSDLDYIVNQPYKYGFTTSVESEQFPRGISREVV
                     KLISKKKNEPEYLLNFRLKAYEKWTKMKNPKWAHLKHPNIDFNSIIYYAVPKLKKELN
                     SLDEVDPEILDTFNKLGISLNEQKRLSNVAVDAVFDSVSIATTFKKELAEAGVIFCSI
                     SEAIRNYPDLIQKYLGTVVPSGDNYFAALNSAVFSDGSFCYIPPDTVCPLELSTYFRI
                     NNEESGQFERTLIVADRGSKVSYLEGCTAPQYDTNQLHAAIVELIALDDAEIKYSTVQ
                     NWYAGNKDGKGGIYNFVTKRGLCSGKNSKISWTQVETGSAITWKYPGCILAGDNSQGE
                     FYSVALTNNYQEADTGTKMIHIGNNTKSKIISKGISAGKSKNSYRGLVKIGPQSFNSR
                     NYSQCDSLLIGQSSQANTFPYIQVQNPTAKVEHEASTSKISEDQIFYFLQRGINLEES
                     VSLMISGFCKDVFNELPMEFAVEADRLLSLKLEGTVG"
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

24:D90812. E.coli genomic DN...[gi:1742768]

Links

```
LOCUS       D90812                 13823 bp    DNA     linear   BCT 29-MAY-1997
DEFINITION  E.coli genomic DNA, Kohara clone #321(38.1-38.4 min.).
ACCESSION   D90812 AB001340
VERSION     D90812.1  GI:1742768
KEYWORDS    Complete and shotgun sequencing; YCF24; ydiD.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
  TITLE     The systematic sequencing of the Escherichia coli genome in Japan
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 13823)
  AUTHORS   Mori,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-DEC-1996) Hirotada Mori, NARA Institute of Science
            and Technology, Res. & Edu. Center for Genetic Info.; 8916-5
            Takayama, Ikoma, Nara 630-01, Japan
            (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
COMMENT     Collaboration Information:
            Project:
             The Japan E.coli genome DNA sequencing project
            Group:
             The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
             Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
             Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
             Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
             Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
             Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
             Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
             Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
             Yamamoto,Y. and Yano,M.
            Headed by:
             Name: Takashi Horiuchi
             Address: National Institute of Basic Biology, Okazaki, 444, Japan
             E-mail: kishori@nibb.ac.jp
            Information operator:
             Name: Hirotada Mori
             Address: NARA Institute of Science and Technology,
                    Ikoma, 630-01, Japan
             E-mail: hmori@gtc.aist-nara.ac.jp
            URL:
             The Japan E. coli genome database
             http:bsw3.aist-nara.ac.jp.
```

APPENDIX D

Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
FEATURES             Location/Qualifiers
     source          1..13823
                     /organism="Escherichia coli"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="38.1 min"
                     /clone="Kohara clone #321"
                     /clone_lib="Kohara lambda miniset library"
                     /note="Nucleotide position 1773491-1787313 from the
                     initiation site of ThrA (0 min.).~This clone is from
                     Kohara lambda miniset library"

...

   gene              888..4416
                     /gene="YCF24"
     CDS             888..2414
                     /gene="YCF24"
                     /note="ORF_ID:o320#13
                     similar to [SwissProt Accession Number P48260]"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA15460.1"
                     /db_xref="GI:1742770"
                     /translation="MWLWRKLWGIGGTMSRNTEATDDVKTWTGGPLNYKEGFFTQLAT
                     DELAKGINEEVVRAISAKRNEPEWMLEFRLNAYRAWLEMEEPHWLKAHYDKLNYQDYS
                     YYSAPSCGNCDDTCASEPGAVQQTGANAFLSKEVEAAFEQLGVPVREGKEVAVDAIFD
                     SVSVATTYREKLAEQGIIFCSFGEAIHDHPELVRKYLGTVVPGNDNFFAALNAAVASD
                     GTFIYVPKGVRCPMELSTYFRINAEKTGQFERTILVADEDSYVSYIEGCSAPVRDSYQ
                     LHAAVVEVIIHKNAEVKYSTVQNWFPGDNNTGGILNFVTKRALCEGENSKMSWTQSET
                     GSAITWKYPSCILRGDNSIGEFYSVALTSGHQQADTGTKMIHIGKNTKSTIISKGISA
                     GHSQNSYRGLVKIMPTATNARNFTQCDSMLIGANCGAHTFPYVECRNNSAQLEHEATT
                     SRIGEDQLFYCLQRGISEEDAISMIVNGFCKDVFSELPLEFAVEAQKLLAISLEHSVG
                     "
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

25:D90811. E.coli genomic DN...[gi:1742754]

Links

```
LOCUS       D90811                19521 bp    DNA     linear   BCT 29-MAY-1997
DEFINITION  E.coli genomic DNA, Kohara clone #320(37.9-38.3 min.).
ACCESSION   D90811 AB001340
VERSION     D90811.1  GI:1742754
KEYWORDS    Complete and shotgun sequencing; YCF24; aroD; ydiB; ydiF.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Inada,T., Isono,K.,
            Itoh,T., Kasai,H., Kashimoto,K., Kimura,S., Kitakawa,M.,
            Kitagawa,M., Makino,K., Miki,T., Mizobuchi,K., Mori,H., Mori,T.,
            Motomura,K., Nakade,S., Nakamura,Y., Nashimoto,H., Nishio,Y.,
            Oshima,T., Saito,N., Sampei,G., Seki,Y., Sivasundaram,S.,
            Tagami,H., Takeda,J., Takemoto,K., Takeuchi,Y., Wada,C.,
            Yamamoto,Y. and Horiuchi,T.
  TITLE     A 570-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 28.0-40.1 min region on the linkage map
  JOURNAL   DNA Res. 3 (6), 363-377 (1996)
  MEDLINE   97251357
   PUBMED   9097039
REFERENCE   2  (sites)
  AUTHORS   Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
            Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
            Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
            Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
            Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
            Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
            Yamamoto,Y. and Yano,M.
  TITLE     The systematic sequencing of the Escherichia coli genome in Japan
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 19521)
  AUTHORS   Mori,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-DEC-1996) Hirotada Mori, NARA Institute of Science
            and Technology, Res. & Edu. Center for Genetic Info.; 8916-5
            Takayama, Ikoma, Nara 630-01, Japan
            (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
COMMENT     Collaboration Information:
            Project:
             The Japan E.coli genome DNA sequencing project
            Group:
             The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
             Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
             Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
             Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
             Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
             Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
             Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N.,
             Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
             Yamamoto,Y. and Yano,M.
```

APPENDIX D

Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
            Headed by:
             Name: Takashi Horiuchi
             Address: National Institute of Basic Biology, Okazaki, 444, Japan
             E-mail: kishori@nibb.ac.jp
            Information operator:
             Name: Hirotada Mori
             Address: NARA Institute of Science and Technology,
                     Ikoma, 630-01, Japan
             E-mail: hmori@gtc.aist-nara.ac.jp
            URL:
             The Japan E. coli genome database
             http:bsw3.aist-nara.ac.jp.
FEATURES             Location/Qualifiers
     source          1..19521
                     /organism="Escherichia coli"
                     /strain="K12"
                     /db_xref="taxon:562"
                     /map="37.9 min"
                     /clone="Kohara clone #320"
                     /clone_lib="Kohara lambda miniset library"
                     /note="Nucleotide position 1760879-1780399 from the
                     initiation site of ThrA (0 min.).~This clone is from
                     Kohara lambda miniset library"

...

     gene            13500..17028
                     /gene="YCF24"
     CDS             13500..15026
                     /gene="YCF24"
                     /note="ORF_ID:o320#13
                     similar to [SwissProt Accession Number P48260]"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA15454.1"
                     /db_xref="GI:1742763"
                     /translation="MWLWRKLWGIGGTMSRNTEATDDVKTWTGGPLNYKEGFFTQLAT
                     DELAKGINEEVVRAISAKRNEPEWMLEFRLNAYRAWLEMEEPHWLKAHYDKLNYQDYS
                     YYSAPSCGNCDDTCASEPGAVQQTGANAFLSKEVEAAFEQLGVPVREGKEVAVDAIFD
                     SVSVATTYREKLAEQGIIFCSFGEAIHDHPELVRKYLGTVVPGNDNFFAALNAAVASD
                     GTFIYVPKGVRCPMELSTYFRINAEKTGQFERTILVADEDSYVSYIEGCSAPVRDSYQ
                     LHAAVVEVIIHKNAEVKYSTVQNWFPGDNNTGGILNFVTKRALCEGENSKMSWTQSET
                     GSAITWKYPSCILRGDNSIGEFYSVALTSGHQQADTGTKMIHIGKNTKSTIISKGISA
                     GHSQNSYRGLVKIMPTATNARNFTQCDSMLIGANCGAHTFPYVECRNNSAQLEHEATT
                     SRIGEDQLFYCLQRGISEEDAISMIVNGFCKDVFSELPLEFAVEAQKLLAISLEHSVG
                     "
```

APPENDIX D
Excerpts from 26 *ycf*24 gene product and gene sequences retrieved from NCBI database

```
26:U30821. Cyanophora parado...[gi:1016083]

Links


LOCUS       CPU30821              135599 bp    DNA     circular PLN 13-NOV-1995
DEFINITION  Cyanophora paradoxa cyanelle, complete genome.
ACCESSION   U30821
VERSION     U30821.1  GI:1016083
KEYWORDS    .
SOURCE      cyanelle Cyanophora paradoxa
  ORGANISM  Cyanophora paradoxa
            Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
REFERENCE   1  (bases 1 to 135599)
  AUTHORS   Stirewalt,V.L., Michalowski,C.B., Luffelhardt,W., Bohnert,H.J. and
            Bryant,D.A.
  TITLE     Nucleotide sequence of the cyanelle genome from Cyanophora paradoxa
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 135599)
  AUTHORS   Bryant,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1995) Donald A. Bryant, Biochemistry and
            Molecular Biology, The Pennsylvania State University, S-234 Frear
            Bldg., University Park, PA 16802, USA
FEATURES             Location/Qualifiers
     source          1..135599
                     /organism="Cyanophora paradoxa"
                     /organelle="plastid:cyanelle"
                     /strain="Pringsheim LB555"
                     /db_xref="taxon:2762"
...

   gene              complement(67893..69353)
                     /gene="ycf24"
     CDS             complement(67893..69353)
                     /gene="ycf24"
                     /codon_start=1
                     /product="ABC transporter subunit"
                     /protein_id="AAA81250.1"
                     /db_xref="GI:1016163"
                     /translation="MVNTQSPKNSGLENLVNQPYKYGLPLIFEIETISKGLTEETIRL
                     ISEKKNEPQFMLEFRLQAYRKWLEMSNEPEWAHLNYPKINYQDMVYYSAPKQKKKLQS
                     LDEVDPTLLETFEKLGIPLTEQKRLANVAVDAIFDSVSVATTFKEELAKEGVIFCPIS
                     EAVQKYPDLIKKYLGSVVSTSDNYFSCLNAAVFSDGSFCYIPKNVRCPLELSTYFRIN
                     NGESGQFERTLIVADEGSYVSYLEGCTAPQFDTNQLHAAVVELVALDNAEIKYSTVQN
                     WYAGDENGKGGIYNFVTKRGLCAGKNSKISWTQVETGSAITWKYPSCVLLGDNSIGEF
                     YSVALTNRYQQADTGTKMIHIGKNTRSRIISKGISAGHSQNSYRGLVKIGPKAVGARN
                     YSQCDSLLIGDNSQANTFPHLQIKNPTAKVEHEASTSKIGEEQIFYFLQRGINAEEAI
                     SLIISGFCREVFNNLPMEFALEADKLLGLKLEGSVG"
```

Appendix E
66 FR 1099, January 5, 2001
"Guidelines for Examination of Patent Application Under 35 U.S.C. 12, ¶1, "Written Description Requirement"

an asserted utility, unless countervailing evidence can be provided that shows that one of ordinary skill in the art would have a legitimate basis to doubt the credibility of such a statement. Similarly, Office personnel must accept an opinion from a qualified expert that is based upon relevant facts whose accuracy is not being questioned; it is improper to disregard the opinion solely because of a disagreement over the significance or meaning of the facts offered.

Once a *prima facie* showing of no specific and substantial credible utility has been properly established, the applicant bears the burden of rebutting it. The applicant can do this by amending the claims, by providing reasoning or arguments, or by providing evidence in the form of a declaration under 37 CFR 1.132 or a patent or a printed publication that rebuts the basis or logic of the *prima facie* showing. If the applicant responds to the *prima facie* rejection, the Office personnel should review the original disclosure, any evidence relied upon in establishing the *prima facie* showing, any claim amendments, and any new reasoning or evidence provided by the applicant in support of an asserted specific and substantial credible utility. It is essential for Office personnel to recognize, fully consider and respond to each substantive element of any response to a rejection based on lack of utility. Only where the totality of the record continues to show that the asserted utility is not specific, substantial, and credible should a rejection based on lack of utility be maintained.

If the applicant satisfactorily rebuts a *prima facie* rejection based on lack of utility under § 101, withdraw the § 101 rejection and the corresponding rejection imposed under § 112, first paragraph.

Dated: December 29, 2000.

Q. Todd Dickinson,

Under Secretary of Commerce for Intellectual Property and Director of the United States Patent and Trademark Office.

[FR Doc. 01-322 Filed 1-4-01; 8:45 am]

BILLING CODE 3510-16-U

DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

[Docket No. 991027288-0264-02]

RIN 0651-AB10

Guidelines for Examination of Patent Applications Under the 35 U.S.C. 112, ¶ 1, "Written Description" Requirement

AGENCY: United States Patent and Trademark Office, Commerce.

ACTION: Notice.

SUMMARY: These Guidelines will be used by USPTO personnel in their review of patent applications for compliance with the "written description" requirement of 35 U.S.C. 112, ¶ 1. These Guidelines supersede the "Revised Interim Guidelines for Examination of Patent Applications Under the 35 U.S.C. 112, ¶ 1 'Written Description' Requirement" that were published in the **Federal Register** at 64 FR 71427, Dec. 21, 1999, and in the Official Gazette at 1231 O.G. 123, Feb. 29, 2000. These Guidelines reflect the current understanding of the USPTO regarding the written description requirement of 35 U.S.C. 112, ¶ 1, and are applicable to all technologies.

DATES: The Guidelines are effective as of January 5, 2001.

FOR FURTHER INFORMATION CONTACT: Stephen Walsh by telephone at (703) 305-9035, by facsimile at (703) 305-9373, by mail to his attention addressed to United States Patent and Trademark Office, Box 8, Washington, DC 20231, or by electronic mail at "stephen.walsh@uspto.gov"; or Linda Therkorn by telephone at (703) 305-8800, by facsimile at (703) 305-8825, by mail addressed to Box Comments, Commissioner for Patents, Washington, DC 20231, or by electronic mail at "linda.therkorn@uspto.gov."

SUPPLEMENTARY INFORMATION: As of the publication date of this notice, these Guidelines will be used by USPTO personnel in their review of patent applications for compliance with the "written description" requirement of 35 U.S.C. 112, ¶ 1. Because these Guidelines only govern internal practices, they are exempt from notice and comment rulemaking under 5 U.S.C. 553(b)(A).

Discussion of Public Comments

Comments were received from 48 individuals and 18 organizations in response to the request for comments on the "Revised Interim Guidelines for Examination of Patent Applications Under the 35 U.S.C. 112, ¶ 1 'Written Description' Requirement" published in the **Federal Register** at 64 FR 71427, Dec. 21, 1999, and in the Official Gazette at 1231 O.G. 123, Feb. 29, 2000. The written comments have been carefully considered.

Overview of Comments

The majority of comments favored issuance of final written description guidelines with minor revisions. Comments pertaining to the written description guidelines are addressed in detail below. A few comments addressed particular concerns with respect to the associated examiner training materials that are available for public inspection at the USPTO web site (www.uspto.gov). Such comments will be taken under advisement in the revision of the training materials; consequently, these comments are not specifically addressed below as they do not impact the content of the Guidelines. Several comments raised issues pertaining to the patentability of ESTs, genes, or genomic inventions with respect to subject matter eligibility (35 U.S.C. 101), novelty (35 U.S.C. 102), or obviousness (35 U.S.C. 103). As these comments do not pertain to the written description requirement under 35 U.S.C. 112, they have not been addressed. However, the aforementioned comments are fully addressed in the "Discussion of Public Comments" in the "Utility Examination Guidelines" Final Notice, which will be published at or about the same time as the present Guidelines.

Responses to Specific Comments

(1) *Comment:* One comment stated that the Guidelines instruct the patent examiner to determine the correspondence between what applicant has described as the essential identifying characteristic features of the invention and what applicant has claimed, and that such analysis will lead to error. According to the comment, the examiner may decide what applicant should have claimed and reject the claim for failure to claim what the examiner considers to be the invention. Another comment suggested that the Guidelines should clarify what is meant by "essential features of the invention." Another comment suggested that what applicant has identified as the "essential distinguishing characteristics" of the invention should be understood in terms of *Fiers* v. *Revel*, 984 F.2d 1164, 1169, 25 USPQ2d 1601, 1605 (Fed. Cir. 1993) ("Conception of a substance claimed *per se* without reference to a process requires conception of its structure, name,

formula, or definitive chemical or physical properties.").

Response: The suggestions have been adopted in part. The purpose of the written description analysis is to confirm that applicant had possession of what is claimed. The Guidelines have been modified to instruct the examiners to compare the scope of the invention claimed with the scope of what applicant has defined in the description of the invention. That is, the Guidelines instruct the examiner to look for consistency between a claim and what provides adequate factual support for the claim as judged by one of ordinary skill in the art from reading the corresponding written description.

(2) *Comment:* Two comments urge that *Regents of the University of California* v. *Eli Lilly & Co.*, 119 F.3d 1559, 43 USPQ2d 1398 (Fed. Cir. 1997), is bad law and should not be followed by the USPTO because it conflicts with binding precedent, such as *Vas-Cath* v. *Mahurkar*, 935 F.2d 1555, 19 USPQ2d 1111 (Fed. Cir. 1991). *Response:* The final Guidelines are based on the Office's current understanding of the law and are believed to be fully consistent with binding precedent of the U.S. Supreme Court and the U.S. Court of Appeals for the Federal Circuit. *Eli Lilly* is a precedential decision by the Court that has exclusive jurisdiction over appeals involving patent law. Accordingly, the USPTO must follow *Eli Lilly*. Furthermore, the USPTO does not view *Eli Lilly* as conflicting with *Vas-Cath*. *Vas-Cath* explains that the purpose of the written description requirement is to ensure that the applicant has conveyed to those of skill in the art that he or she was in possession of the claimed invention at the time of filing. *Vas-Cath*, 935 F.2d at 1563–64, 19 USPQ2d at 1117. *Eli Lilly* explains that a chemical compound's name does not necessarily convey a written description of the named chemical compound, particularly when a genus of compounds is claimed. *Eli Lilly*, 119 F.3d at 1568, 43 USPQ2d at 1405. The name, if it does no more than distinguish the claimed genus from all others by function, does not satisfy the written description requirement because "it does not define any structural features commonly possessed by members of the genus that distinguish them from others. One skilled in the art therefore cannot, as one can do with a fully described genus, visualize or recognize the identity of the members of the genus." *Eli Lilly*, 119 F.3d at 1568, 43 USPQ2d at 1406. Thus, *Eli Lilly* identified a set of circumstances in which the words of the claim did not, without more, adequately convey to others that applicants had possession of what they claimed.

(3) *Comment:* Several comments urged that the Guidelines do not recognize the inconsistency between the original claim doctrine and the written description requirement as set out in *Fiers* and *Eli Lilly*. On the other hand, another comment asserts that there is no strong presumption that an originally filed claim constitutes an adequate written description of the claimed subject matter. Several comments indicate that *in haec verba* support should be sufficient to comply with the written description requirement. Two comments urge that the concept of constructive reduction to practice upon filing of an application has been ignored. *Response:* As noted above, the USPTO does not find *Fiers* and *Eli Lilly* to be in conflict with binding precedent. An original claim may provide written description for itself, but it still must be an *adequate* written description which establishes that the inventor was in possession of the invention. The "original claim doctrine" is founded on cases which stand for the proposition that originally filed claims are part of the written description of an application as filed, and thus subject matter which is present only in originally filed claims need not find independent support in the specification. *See, e.g., In re Koller*, 613 F.2d 819, 824, 204 USPQ 702, 706 (CCPA 1980) (later added claims of similar scope and wording were adequately described by original claims); *In re Gardner*, 480 F.2d 879, 880, 178 USPQ 149, 149 (CCPA 1973) ("Under these circumstances, we consider the original claim in itself adequate 'written description' of the claimed invention. It was equally a 'written description' * * * whether located among the original claims or in the descriptive part of the specification."). However, as noted in the preceding comment, *Eli Lilly* identified a set of circumstances in which the words of the claim did not, without more, adequately convey to others that applicants had possession of what they claimed. When the name of a novel chemical compound does not convey sufficient structural information about the compound to identify the compound, merely reciting the name is not enough to show that the inventor had possession of the compound at the time the name was written. The Guidelines indicate that there is a "strong presumption" that an adequate written description of the claimed invention is present when the application is filed, consistent with *In re Wertheim*, 541 F.2d 257, 263, 191 USPQ 90, 97 (CCPA 1976) ("we are of the opinion that the PTO has the initial burden of presenting evidence or reasons why persons skilled in the art would not recognize in the disclosure a description of the invention defined by the claims."). In most cases, the statement that "an originally filed claim is its own written description," is borne out because the claim language conveys to others of skill in the art that the applicant was "in possession" of what is claimed. The Guidelines emphasize that the burden of proof is on the examiner to establish that a description as filed is not adequate and require the examiner to introduce sufficient evidence or technical reasoning to shift the burden of going forward with contrary evidence to the applicant.

(4) *Comment:* One comment stated that the Guidelines change the substance of the written description requirement to require some level of enablement. The comment stated that the *Eli Lilly* case should not be followed because its change in the quality of the description required is in conflict with precedent. Another comment suggested that to comply with the written description requirement, the description must both (i) demonstrate possession of the claimed invention by the applicant; and (ii) put the public in possession of the claimed invention. *Response:* As noted in the comment above, the USPTO is bound by the Federal Circuit's decision in *Eli Lilly*. The Guidelines have been revised to clarify that an applicant must provide a description of the claimed invention which shows that applicant was in possession of the claimed invention. The suggestion to emphasize that the written description requirement must put the public in possession of the invention has not been adopted because it removes much of the distinction between the written description requirement and the enablement requirement. Although the two concepts are entwined, they are distinct and each is evaluated under separate legal criteria. The written description requirement, a question of fact, ensures that the inventor conveys to others that he or she had possession of the claimed invention; whereas, the enablement requirement, a question of law, ensures that the inventor conveys to others how to make and use the claimed invention.

(5) *Comment:* One comment suggested that the Guidelines should provide examples of situations in which the written description requirement was met but the enablement requirement was not, and vice versa. Another comment stated that examiners often use enablement language in making

written description rejections.

Response: The enablement and written description requirements are not coextensive and, therefore, situations will arise in which one requirement is met but the other is not. Federal Circuit case law demonstrates many circumstances where enablement or written description issues, but not both, were before the Court. These Guidelines are intended to clarify for the examining corps the criteria needed to satisfy the written description requirement. For examples applying these Guidelines to hypothetical fact situations, see the "Synopsis of Application of Written Description Guidelines" (examiner training materials available on-line at *http://www.uspto.gov/web/menu/written.pdf).* These examples, as well as the examination form paragraphs and instructions on their proper use, provide the appropriate language examiners should use in making written description rejections.

(6) *Comment:* One comment disagreed with the statement in an endnote that "the fact that a great deal more than just a process is necessary to render a product invention obvious means that a great deal more than just a process is necessary to provide written description for a product invention." The comment indicated that the statement is overly broad and inconsistent with the "strong presumption that an adequate written description of the claimed invention is present when the application is filed." As an extreme case, for example, for product-by-process claims, nothing else would be needed to provide the written description of the product. *Response:* The endnote has been clarified and is now more narrowly drawn. However, there is no *per se* rule that disclosure of a process is sufficient to adequately describe the products produced by the process. In fact, *Fiers* v. *Revel* and *Eli Lilly* involved special circumstances where the disclosure of a process of making and the function of the product alone did not provide an adequate written description for product claims. Even when a product is claimed in a product-by-process format, the adequacy of the written description of the process to support product claims must be evaluated on a case-by-case basis.

(7) *Comment:* Several comments urge that actual reduction to practice, as a method of satisfying the written description requirement by demonstrating possession, has been over-emphasized. *Response:* The Guidelines have been clarified to state that describing an actual reduction to practice is one of a number of ways to show possession of the invention. Description of an actual reduction to practice offers an important "safe haven" that applies to all applications and is just one of several ways by which an applicant may demonstrate possession of the claimed invention. Actual reduction to practice may be crucial in the relatively rare instances where the level of knowledge and level of skill are such that those of skill in the art cannot describe a composition structurally, or specify a process of making a composition by naming components and combining steps, in such a way as to distinguish the composition with particularity from all others. Thus, the emphasis on actual reduction to practice is appropriate in those cases where the inventor cannot provide an adequate description of what the composition is, and a definition by function is insufficient to define a composition "because it is only an indication of what the [composition] does, rather than what it is." *Eli Lilly*, 119 F.3d at 1568, 43 USPQ at 1406. *See also Amgen Inc.* v. *Chugai Pharmaceutical Co.*, 927 F.2d 1200, 1206, 18 USPQ2d 1016, 1021 (Fed. Cir. 1991).

(8) *Comment:* One comment asserts that the citation to *Pfaff* v. *Wells Electronics, Inc.*, 525 U.S. 55, 48 USPQ2d 1641 (1998) is inappropriate and should be deleted because *Pfaff* is concerned with § 102(b) on-sale bar, not written description. Another comment suggested that the Guidelines should provide an explanation of how the "ready for patenting" concept of *Pfaff* should be used in determining compliance with the written description requirement. *Response:* The Guidelines state the general principle that actual reduction to practice is not required to show possession of, or to adequately describe, a claimed invention (although, as noted in the previous comment, an actual reduction to practice is crucial in relatively rare instances). An alternative is to show that the invention described was "ready for patenting" as set out in *Pfaff*. For example, a *description* of activities that demonstrates the invention was "ready for patenting" satisfies the written description requirement. As *Wertheim* indicates, "how the specification accomplishes this is not material." 541 F.2d at 262, 191 USPQ at 96.

(9) *Comment:* One comment stated that the written description of a claimed DNA should be required to include the complete sequence of the DNA and claims should be limited to the DNA sequence disclosed. *Response:* Describing the complete chemical structure, *i.e.*, the DNA sequence, of a claimed DNA is one method of satisfying the written description requirement, but it is not the only method. *See Eli Lilly*, 119 F.3d at 1566, 43 USPQ2d at 1404 ("An adequate written description of a DNA * * * requires a precise definition, *such as* by structure, formula, chemical name, or physical properties." (emphasis added, internal quote omitted)). Therefore, there is no basis for a *per se* rule requiring disclosure of complete DNA sequences or limiting DNA claims to only the sequence disclosed.

(10) *Comment:* One comment stated that it is difficult to envision how one could provide a description of sufficient identifying characteristics of the invention without physical possession of a species of the invention, and thus this manner of showing possession should be considered as a way to show actual reduction to practice. *Response:* This suggestion has not been adopted. The three ways of demonstrating possession as set forth in the Guidelines are merely exemplary and are not mutually exclusive. While there are some cases where a description of sufficient relevant identifying characteristics will evidence an actual reduction to practice, there are other cases where it will not. *See, e.g., Ralston Purina Co.* v. *Far-Mar-Co, Inc.*, 772 F.2d 1570, 1576, 227 USPQ 177, 180 (Fed. Cir. 1985) (disclosure taken with the knowledge of those skilled in the art may be sufficient support for claims).

(11) *Comment:* One comment stated that the Guidelines should be revised to indicate that the test of disclosure of sufficiently detailed drawings should be expanded to include structural claiming of chemical entities. *Response:* The suggestion has been adopted.

(12) *Comment:* One comment stated that the Guidelines should reflect that an inventor is in possession of the invention when the inventor demonstrably has at least a complete conception thereof, and that factors and attributes which provide proof of written description should include evidence typically provided to prove a complete conception. *Response:* The suggestion has not been adopted because the conception analysis typically involves documentary evidence in addition to the description of the invention in the application as filed. However, it is acknowledged that if evidence typically provided to prove a complete conception is present in the specification as filed, it would be sufficient to show possession. The Federal Circuit has stated "[t]he conception analysis necessarily turns on the inventor's ability to describe his invention with particularity. Until he can do so, he cannot prove possession

of the complete mental picture of the invention." *Burroughs Wellcome Co. v. Barr Labs., Inc.*, 40 F.3d 1223, 1228, 32 USPQ2d 1915, 1919 (Fed. Cir. 1994). As further noted by the Federal Circuit, in order to prove conception, "a party must show possession of every feature recited in the count, and that every limitation of the count must have been known to the inventor at the time of the alleged conception." *Coleman v. Dines*, 754 F.2d 353, 359, 224 USPQ 857, 862 (Fed. Cir. 1985).

(13) *Comment:* One comment indicated that a "possession" test does not appear in Title 35 of the U.S. Code and is not clearly stated by the Federal Circuit. Therefore, it is recommended that patent examiners be directed to use existing judicial precedent to make rejections of claims unsupported by a statutory written description requirement. *Response:* While the Federal Circuit has not specifically laid out a "possession" test, the Court has clearly indicated that possession is a cornerstone of the written description inquiry. *See, e.g., Vas-Cath, Inc. v. Mahurkar*, 935 F.2d 1555, 1563, 19 USPQ2d 1111, 1116 (Fed. Cir. 1991); *see also Purdue Pharma L.P. v. Faulding Inc.*, 230 F.3d 1320, 1323, 56 USPQ2d 1481, 1483 (Fed. Cir. 2000) ("[o]ne skilled in the art, reading the disclosure, must immediately discern the limitation at issue in the claims") (internal quote omitted). The possession test as set forth in the Guidelines is extrapolated from case law in a wide variety of technologies and is not intended to be limiting. Any rejections made by examiners will be made under 35 U.S.C. 112, ¶1, with supporting rationale. Final rejections are appealable if applicant disagrees and follows the required procedures to appeal.

(14) *Comment:* Two comments indicated that if the amino acid sequence for a polypeptide whose utility has been identified is described, then the question of possession of a class of nucleotides encoding that polypeptide can be addressed as a relatively routine matter using the understanding of the genetic code, and that the endnote addressing this issue should be revised. *Response:* The suggestion of these comments has been incorporated in the Guidelines and will be reflected in the training materials. However, based upon *In re Bell*, 991 F.2d 781, 785, 26 USPQ2d 1529, 1532 (Fed. Cir. 1993) and *In re Baird*, 16 F.3d 380, 382, 29 USPQ2d 1550, 1552 (Fed. Cir. 1994), this does not mean that applicant was in possession of any particular species of the broad genus.

(15) *Comment:* One comment disagreed with an endnote which stated that a laundry list disclosure of moieties does not constitute a written description of every species in a genus. Specifically, the comment indicates that if the existence of a functional genus is adequately described in the specification, a laundry list of the species within that genus must satisfy the written description requirement. *Response:* The suggestion to revise the endnote will not be adopted. A lack of adequate written description problem arises if the knowledge and level of skill in the art would not permit one skilled in the art to immediately envisage the product claimed from the disclosure. This was aptly demonstrated in *In re Bell* and *In re Baird* where possession of a large genus did not put a person of ordinary skill in the art in possession of any particular species. *See also Purdue Pharma*, 230 F.3d at 1328, 56 USPQ2d at 1487 (because the original specification did not disclose the later claimed concentration ratio was a part of the invention, the inventors cannot argue that they are merely narrowing a broad invention).

(16) *Comment:* One comment suggested that in the majority of cases, a single species will support a generic claim, and that the Guidelines should emphasize this point. *Response:* The suggestion has been adopted to a limited degree. The Guidelines now indicate that a single species may, in some instances, provide an adequate written description of a generic claim when the description of the species would evidence to one of ordinary skill in the art that the invention includes the genus. Note, however, *Tronzo v. Biomet, Inc.*, 156 F.3d 1154, 47 USPQ2d 1829 (Fed. Cir. 1998), where the species in the parent application was held not to provide written description support for the genus in the child application.

(17) *Comment:* One comment asserted that the Guidelines should focus on the compliance of the claims, not the specification, with the written description requirement. *Response:* This suggestion will not be adopted. "The specification shall contain a written description of the invention." 35 U.S.C. 112. The claims are part of the specification. *Id.*, ¶ 2. If an adequate description is provided, it will suffice "whether located among the original claims or in the descriptive part of the specification." *In re Gardner*, 480 F.2d 879, 880, 178 USPQ 149 (CCPA 1973). The entire disclosure, including the specification, drawings, and claims, must be considered.

(18) *Comment:* One comment asserted that the Guidelines confuse "new matter," 35 U.S.C. 132, with the written description requirement, and that the same standard for written description should be applied to both original claims and new or amended claims. *Response:* The Guidelines indicate that for both original and amended claims, the inquiry is whether one skilled in the art can reasonably conclude that the inventor had possession of the claimed invention at the time the application was filed.

(19) *Comment:* One comment suggested that the second paragraph of the section pertaining to determining what the claim as a whole covers should be deleted because it relates more to compliance with § 112, second paragraph, than with the written description requirement. *Response:* This suggestion will not be adopted. The claims must be construed and all issues as to the scope and meaning of the claim must be explored during the inquiry into whether the written description requirement has been met. The concept of treating the claim as a whole is applicable to all criteria for patentability.

(20) *Comment:* One comment suggested a different order for the general analysis for determining compliance with the written description requirement, starting with reading the claim, then the specification, and then determining whether the disclosure demonstrates possession by the applicant. *Response:* This suggestion will not be adopted. The claims must be construed as broadly as reasonable in light of the specification and the knowledge in the art. *See In re Morris*, 127 F.3d 1048, 1054, 44 USPQ2d 1023, 1027 (Fed. Cir. 1997). Then the disclosure must be evaluated to determine whether it adequately describes the claimed invention, i.e., whether it conveys to a person having ordinary skill in the art that the applicant had possession of what he or she now claims.

(21) *Comment:* Several comments suggested that the Guidelines are unclear with regard to how the examiner should treat the transitional phrase "consisting essentially of." The comments also suggested that the endnote that explains "consisting essentially of" does not make clear how the use of this intermediate transitional language affects the scope of the claim. Several comments stated that the USPTO does not have legal authority to treat claims reciting this language as open (equivalent to "comprising"). Another comment suggested that the phrase "clear indication in the specification" be replaced with "explicit or implicit indication." *Response:* The transitional phrase "consisting essentially of" "excludes

ingredients that would 'materially affect the basic and novel characteristics' of the claimed composition." *Atlas Powder Co.* v. *E.I. DuPont de Nemours & Co.*, 750 F.2d 1569, 1574, 224 USPQ 409, 412 (Fed. Cir. 1984). The basic and novel characteristics of the claimed invention are limited by the balance of the claim. *In re Janakirama-Rao*, 317 F.2d 951, 954, 137 USPQ 893, 896 (CCPA 1963). However, during prosecution claims must be read broadly, consistent with the specification. *In re Morris*, 127 F.3d 1048, 1054, 44 USPQ2d 1023, 1027 (Fed. Cir. 1997). Thus, *for purposes of searching for and applying prior art in a rejection under 35 U.S.C. 102 or 103*, if the specification or the claims do not define the "basic and novel" properties of the claimed subject matter (or if such properties are in dispute), the broadest reasonable interpretation consistent with the specification is that the basic and novel characteristics are merely the presence of the recited limitations. *See, e.g., Janakirama-Rao*, 317 F.2d at 954, 137 USPQ at 895–96. This does not indicate that the intermediate transitional language is never given weight. Applicants may amend the claims to avoid the rejections or seek to establish that the specification provides definitions of terms in the claims that define the basic and novel characteristics of the claimed invention which distinguish the claimed invention from the prior art. When an applicant contends that additional steps or materials in the prior art are excluded by the recitation of 'consisting essentially of,' applicant has the burden of showing that the introduction of additional steps or components would materially change the characteristics of applicant's invention. *In re De Lajarte*, 337 F.2d 870, 143 USPQ 256 (CCPA 1964). The language used in the Guidelines is consistent with *PPG Industries Inc.* v. *Guardian Industries Corp.*, 156 F.3d 1351, 1355, 48 USPQ2d 1351, 1355 (Fed. Cir. 1998) ("PPG could have defined the scope of the phrase 'consisting essentially of' for purposes of its patent by making clear in its specification what it regarded as constituting a material change in the basic and novel characteristics.").

(22) *Comment:* One comment stated that the written description should "disclose the invention," including why the invention works and how it was developed. *Response:* This suggestion has not been adopted. An inventor does not need to know how or why the invention works in order to obtain a patent. *Newman* v. *Quigg*, 877 F.2d 1575, 1581, 11 USPQ2d 1340, 1345 (Fed. Cir. 1989). To satisfy the enablement requirement of 35 U.S.C. 112, ¶1, an application must disclose the claimed invention in sufficient detail to enable a person of ordinary skill in the art to make and use the claimed invention. To satisfy the written description requirement of 35 U.S.C. 112, ¶ 1, the description must show that the applicant was in possession of the claimed invention at the time of filing. There is no statutory basis to require disclosure of why an invention works or how it was developed. "Patentability shall not be negatived by the manner in which the invention was made." 35 U.S.C. 103(a).

(23) *Comment:* One comment recommended that the phrases "emerging and unpredictable technologies" and "unpredictable art" be replaced with the phrase—inventions characterized by factors which are not reasonably predictable in terms of the ordinary skill in the art—. *Response:* The suggestion is adopted in part and the recommended phrase has been added as an alternative.

(24) *Comment:* One comment recommended that the phrase "conventional in the art" be replaced with—part of the knowledge of one of ordinary skill in the art—. *Response:* The suggestion is adopted in part and the recommended phrase has been added as an alternative. The standard of "conventional in the art" is supported by case law holding that a patent specification "need not teach, and preferably omits, what is well known in the art." *See Spectra-Physics, Inc.* v. *Coherent, Inc.*, 827 F.2d 1524, 1534, 3 USPQ2d 1737, 1743 (Fed. Cir. 1987); *Hybritech Inc.* v. *Monoclonal Antibodies, Inc.*, 802 F.2d 1367, 1384, 231 USPQ 81, 94 (Fed. Cir. 1986). *See also Atmel Corp.* v. *Information Storage Devices, Inc.*, 198 F.3d 1374, 1382, 53 USPQ2d 1225, 1231 (Fed. Cir. 1999).

(25) *Comment:* One comment recommended that the Guidelines be amended to state that the appropriate skill level for determining possession of the claimed invention is that of a person of *ordinary* skill in the art. *Response:* The comment has not been adopted. The statutory language itself indicates that compliance with the requirements of 35 U.S.C. 112, ¶1, is judged from the standard of "any person skilled in the art." It is noted, however, that the phrases "one of skill in the art" and "one of ordinary skill in the art" appear to be synonymous. *See, e.g., Union Oil Co.* v. *Atlantic Richfield Co.*, 208 F.3d 989, 997, 54 USPQ2d 1227, 1232 (Fed. Cir. 2000) ("The written description requirement does not require the applicant 'to describe exactly the subject matter claimed, [instead] the description must clearly allow *persons of ordinary skill in the art* to recognize that [he or she] invented what is claimed.' Thus, § 112, ¶ 1, ensures that, as of the filing date, the inventor conveyed with reasonable clarity to *those of skill in the art* that he was in possession of the subject matter of the claims." (citations omitted, emphasis added)).

(26) *Comment:* One comment stated that an endnote misstates the relevant law in stating that, to show inherent written descriptive support for a claim limitation, the inherent disclosure must be such as would be recognized by a person of ordinary skill in the art. The comment recommended that the endnote be amended to delete the reference to recognition by persons of ordinary skill and to cite *Pingree* v. *Hull*, 518 F.2d 624, 186 USPQ 248 (CCPA 1975), rather than *In re Robertson*, 169 F.3d 743, 49 USPQ2d 1949 (Fed. Cir. 1999). *Response:* The comment has not been adopted. Federal Circuit precedent makes clear that an inherent disclosure must be recognized by those of ordinary skill in the art. *See, e.g., Hyatt* v. *Boone*, 146 F.3d 1348, 1354–55, 47 USPQ2d 1128, 1132 (Fed. Cir. 1998) ("[T]he purpose of the description requirement is 'to ensure that the inventor had possession, as of the filing date of the application relied on, of the specific subject matter later claimed by him.' * * * Thus, the written description must include all of the limitations of the interference count, or the applicant must show that any absent text is necessarily comprehended in the description provided *and would have been so understood* at the time the patent application was filed." (emphasis added)). *See also Reiffin* v. *Microsoft Corp.*, 214 F.3d 1342, 1346, 54 USPQ2d 1915, 1917 (Fed. Cir. 2000) (The "application considered as a whole must convey to one of ordinary skill in the art, either explicitly or inherently, that [the inventor] invented the subject matter claimed * * *. See * * * *Continental Can Co. USA* v. *Monsanto Co.*, 948 F.2d 1264, 1268, 20 USPQ2d 1746, 1749 (Fed. Cir. 1991) (descriptive matter may be inherently present in a specification if one skilled in the art would necessarily recognize such a disclosure)").

(27) *Comment:* Several comments pointed out an inconsistency in the **Federal Register** Notice re: the Revised Interim Written Description Guidelines. The inconsistency concerned the treatment of claims directed to an isolated DNA comprising SEQ ID NO:1 wherein SEQ ID NO:1 is an expressed sequence tag. The comments contrasted paragraphs 34 and 35 of the Response to

Public Comments with the statement in the text of the Guidelines that a genus must be supported by a representative number of species (as analyzed in Example 7 of the training materials). *Response:* The USPTO acknowledges that there was an inconsistency. The Office notes that a claim reciting a nucleic acid comprising SEQ ID NO:1 may be subject to a rejection for lack of an adequate written description where particular identifiable species within the scope of the claim lack an adequate written description. The training materials as amended exemplify an appropriate analysis.

(28) *Comment:* One comment stated that the USPTO should respond to the issue of whether the U.S. is meeting its TRIPs obligations. This comment noted that the USPTO did not address an earlier comment regarding the "Interim Guidelines for the Examination of Patent Applications under the 35 U.S.C. 112, ¶ 1, 'Written Description' Requirement," 63 FR 32,639, June 15, 1998, which questioned whether the written description requirement is truly different from the enablement requirement, and indicated that such a requirement may be contrary to the TRIPs provisions of the World Trade Organization (Article 27.1). Article 27.1 requires WTO Members to, *inter alia*, make patents available, with limited exceptions, for products and processes in all fields of technology so long as those products and processes are new, involve an inventive step, and are capable of industrial application. The comment further suggested a response. *Response:* TRIPs Article 27 does not address what must be included in a patent application to allow WTO Member officials to determine whether particular inventions meet the standards for patentability established in that Article. TRIPs Article 29, which is more relevant to this comment, states that Members "shall require" patent applicants to disclose their invention "in a manner sufficiently clear and complete for the invention to be carried out by a person skilled in the art." If the written description is not clear and complete, the applicant may not have been in possession of the invention. This may support both written description and enablement standards. In addition, Article 29 expressly authorizes Members to require patent applicants to disclose the best method the inventor knows at the time of filing an application for carrying out the invention.

(29) *Comment:* Two comments commended the USPTO for eliminating the Biotechnology Specific Examples in the Revised Interim Written Description Guidelines and providing separate training materials. One comment indicated a need to reconfirm the examples set forth in the Interim Written Description Guidelines published in 1998. *Response:* The current training materials reflect the manner in which the USPTO interprets the Written Description Guidelines.

(30) *Comment:* Several comments addressed specific concerns about the examiner training materials. *Response:* The comments received with respect to the training materials will be taken under advisement as the Office revises the training materials in view of the revisions to the Guidelines. The specific comments will not be addressed herein as they do not impact the language of the Guidelines.

Guidelines for the Examination of Patent Applications Under the 35 U.S.C. 112, ¶ 1, "Written Description" Requirement

These "Written Description Guidelines" are intended to assist Office personnel in the examination of patent applications for compliance with the written description requirement of 35 U.S.C. 112, ¶ 1. This revision is based on the Office's current understanding of the law and public comments received in response to the USPTO's previous request for public comments on its Revised Interim Written Description Guidelines and is believed to be fully consistent with binding precedent of the U.S. Supreme Court, as well as the U.S. Court of Appeals for the Federal Circuit and its predecessor courts.

This revision does not constitute substantive rulemaking and hence does not have the force and effect of law. It is designed to assist Office personnel in analyzing claimed subject matter for compliance with substantive law. Rejections will be based upon the substantive law, and it is these rejections which are appealable. Consequently, any perceived failure by Office personnel to follow these Guidelines is neither appealable nor petitionable.

These Guidelines are intended to form part of the normal examination process. Thus, where Office personnel establish a *prima facie* case of lack of written description for a claim, a thorough review of the prior art and examination on the merits for compliance with the other statutory requirements, including those of 35 U.S.C. 101, 102, 103, and 112, is to be conducted prior to completing an Office action which includes a rejection for lack of written description. Office personnel are to rely on this revision of the Guidelines in the event of any inconsistent treatment of issues involving the written description requirement between these Guidelines and any earlier guidance provided from the Office.

I. General Principles Governing Compliance With the "Written Description" Requirement for Applications

The first paragraph of 35 U.S.C. 112 requires that the "specification shall contain a written description of the invention * * *." This requirement is separate and distinct from the enablement requirement.[1] The written description requirement has several policy objectives. "[T]he 'essential goal' of the description of the invention requirement is to clearly convey the information that an applicant has invented the subject matter which is claimed."[2] Another objective is to put the public in possession of what the applicant claims as the invention.[3] The written description requirement of the Patent Act promotes the progress of the useful arts by ensuring that patentees adequately describe their inventions in their patent specifications in exchange for the right to exclude others from practicing the invention for the duration of the patent's term.

To satisfy the written description requirement, a patent specification must describe the claimed invention in sufficient detail that one skilled in the art can reasonably conclude that the inventor had possession of the claimed invention.[4] An applicant shows possession of the claimed invention by describing the claimed invention with all of its limitations using such descriptive means as words, structures, figures, diagrams, and formulas that fully set forth the claimed invention.[5] Possession may be shown in a variety of ways including description of an actual reduction to practice,[6] or by showing that the invention was "ready for patenting" such as by the disclosure of drawings or structural chemical formulas that show that the invention was complete,[7] or by describing distinguishing identifying characteristics sufficient to show that the applicant was in possession of the claimed invention.[8] A question as to whether a specification provides an adequate written description may arise in the context of an original claim which is not described sufficiently, a new or amended claim wherein a claim limitation has been added or removed, or a claim to entitlement of an earlier priority date or effective filing date under 35 U.S.C. 119, 120, or 365(c).[9] Compliance with the written description requirement is a question of

fact which must be resolved on a case-by-case basis.[10]

A. Original Claims

There is a strong presumption that an adequate written description of the claimed invention is present when the application is filed.[11] However, the issue of a lack of adequate written description may arise even for an original claim when an aspect of the claimed invention has not been described with sufficient particularity such that one skilled in the art would recognize that the applicant had possession of the claimed invention.[12] The claimed invention as a whole may not be adequately described if the claims require an essential or critical feature which is not adequately described in the specification and which is not conventional in the art or known to one of ordinary skill in the art.[13] This problem may arise where an invention is described solely in terms of a method of its making coupled with its function and there is no described or art-recognized correlation or relationship between the structure of the invention and its function.[14] A lack of adequate written description issue also arises if the knowledge and level of skill in the art would not permit one skilled in the art to immediately envisage the product claimed from the disclosed process.[15]

B. New or Amended Claims

The proscription against the introduction of new matter in a patent application [16] serves to prevent an applicant from adding information that goes beyond the subject matter originally filed.[17] Thus, the written description requirement prevents an applicant from claiming subject matter that was not adequately described in the specification as filed. New or amended claims which introduce elements or limitations which are not supported by the as-filed disclosure violate the written description requirement.[18] While there is no *in haec verba* requirement, newly added claim limitations must be supported in the specification through express, implicit, or inherent disclosure. An amendment to correct an obvious error does not constitute new matter where one skilled in the art would not only recognize the existence of the error in the specification, but also recognize the appropriate correction.[19] Deposits made after the application filing date cannot be relied upon to support additions to or correction of information in the application as filed.[20]

Under certain circumstances, omission of a limitation can raise an issue regarding whether the inventor had possession of a broader, more generic invention.[21] A claim that omits an element which applicant describes as an essential or critical feature of the invention originally disclosed does not comply with the written description requirement.[22]

The fundamental factual inquiry is whether the specification conveys with reasonable clarity to those skilled in the art that, as of the filing date sought, applicant was in possession of the invention as now claimed.[23]

II. Methodology for Determining Adequacy of Written Description

A. Read and Analyze the Specification for Compliance With 35 U.S.C. 112, ¶ 1

Office personnel should adhere to the following procedures when reviewing patent applications for compliance with the written description requirement of 35 U.S.C. 112, ¶ 1. The examiner has the initial burden, after a thorough reading and evaluation of the content of the application, of presenting evidence or reasons why a person skilled in the art would not recognize that the written description of the invention provides support for the claims. There is a strong presumption that an adequate written description of the claimed invention is present in the specification as filed;[24] however, with respect to newly added or amended claims, applicant should show support in the original disclosure for the new or amended claims.[25] Consequently, rejection of an original claim for lack of written description should be rare. The inquiry into whether the description requirement is met is a question of fact that must be determined on a case-by-case basis.[26]

1. For Each Claim, Determine What the Claim as a Whole Covers

Claim construction is an essential part of the examination process. Each claim must be separately analyzed and given its broadest reasonable interpretation in light of and consistent with the written description.[27] The entire claim must be considered, including the preamble language [28] and the transitional phrase.[29] The claim as a whole, including all limitations found in the preamble,[30] the transitional phrase, and the body of the claim, must be sufficiently supported to satisfy the written description requirement.[31]

The examiner should evaluate each claim to determine if sufficient structures, acts, or functions are recited to make clear the scope and meaning of the claim, including the weight to be given the preamble.[32] The absence of definitions or details for well-established terms or procedures should not be the basis of a rejection under 35 U.S.C. 112, ¶ 1, for lack of adequate written description. Limitations may not, however, be imported into the claims from the specification.

2. Review the Entire Application to Understand How Applicant Provides Support for the Claimed Invention Including Each Element and/or Step

Prior to determining whether the disclosure satisfies the written description requirement for the claimed subject matter, the examiner should review the claims and the entire specification, including the specific embodiments, figures, and sequence listings, to understand how applicant provides support for the various features of the claimed invention.[33] The analysis of whether the specification complies with the written description requirement calls for the examiner to compare the scope of the claim with the scope of the description to determine whether applicant has demonstrated possession of the claimed invention. Such a review is conducted from the standpoint of one of skill in the art at the time the application was filed [34] and should include a determination of the field of the invention and the level of skill and knowledge in the art. Generally, there is an inverse correlation between the level of skill and knowledge in the art and the specificity of disclosure necessary to satisfy the written description requirement. Information which is well known in the art need not be described in detail in the specification.[35]

3. Determine Whether There is Sufficient Written Description to Inform a Skilled Artisan That Applicant was in Possession of the Claimed Invention as a Whole at the Time the Application Was Filed

a. Original claims. Possession may be shown in many ways. For example, possession may be shown, *inter alia*, by describing an actual reduction to practice of the claimed invention. Possession may also be shown by a clear depiction of the invention in detailed drawings or in structural chemical formulas which permit a person skilled in the art to clearly recognize that applicant had possession of the claimed invention. An adequate written description of the invention may be shown by any description of sufficient, relevant, identifying characteristics so long as a person skilled in the art would recognize that the inventor had possession of the claimed invention.[36]

A specification may describe an actual reduction to practice by showing

that the inventor constructed an embodiment or performed a process that met all the limitations of the claim and determined that the invention would work for its intended purpose.[37] Description of an actual reduction to practice of a biological material may be shown by specifically describing a deposit made in accordance with the requirements of 37 CFR 1.801 *et seq.*[38]

An applicant may show possession of an invention by disclosure of drawings[39] or structural chemical formulas[40] that are sufficiently detailed to show that applicant was in possession of the claimed invention as a whole. The description need only describe in detail that which is new or not conventional.[41] This is equally true whether the claimed invention is directed to a product or a process.

An applicant may also show that an invention is complete by disclosure of sufficiently detailed, relevant identifying characteristics[42] which provide evidence that applicant was in possession of the claimed invention,[43] *i.e.*, complete or partial structure, other physical and/or chemical properties, functional characteristics when coupled with a known or disclosed correlation between function and structure, or some combination of such characteristics.[44] What is conventional or well known to one of ordinary skill in the art need not be disclosed in detail.[45] If a skilled artisan would have understood the inventor to be in possession of the claimed invention at the time of filing, even if every nuance of the claims is not explicitly described in the specification, then the adequate description requirement is met.[46]

(1) For each claim drawn to a single embodiment or species:[47]

(a) Determine whether the application describes an actual reduction to practice of the claimed invention.

(b) If the application does not describe an actual reduction to practice, determine whether the invention is complete as evidenced by a reduction to drawings or structural chemical formulas that are sufficiently detailed to show that applicant was in possession of the claimed invention as a whole.

(c) If the application does not describe an actual reduction to practice or reduction to drawings or structural chemical formula as discussed above, determine whether the invention has been set forth in terms of distinguishing identifying characteristics as evidenced by other descriptions of the invention that are sufficiently detailed to show that applicant was in possession of the claimed invention.

(i) Determine whether the application as filed describes the complete structure (or acts of a process) of the claimed invention as a whole. The complete structure of a species or embodiment typically satisfies the requirement that the description be set forth "in such full, clear, concise, and exact terms" to show possession of the claimed invention.[48] If a complete structure is disclosed, the written description requirement is satisfied for that species or embodiment, and a rejection under 35 U.S.C. 112, ¶ 1, for lack of written description must not be made.

(ii) If the application as filed does not disclose the complete structure (or acts of a process) of the claimed invention as a whole, determine whether the specification discloses other relevant identifying characteristics sufficient to describe the claimed invention in such full, clear, concise, and exact terms that a skilled artisan would recognize applicant was in possession of the claimed invention.[49]

Whether the specification shows that applicant was in possession of the claimed invention is not a single, simple determination, but rather is a factual determination reached by considering a number of factors. Factors to be considered in determining whether there is sufficient evidence of possession include the level of skill and knowledge in the art, partial structure, physical and/or chemical properties, functional characteristics alone or coupled with a known or disclosed correlation between structure and function, and the method of making the claimed invention. Disclosure of any combination of such identifying characteristics that distinguish the claimed invention from other materials and would lead one of skill in the art to the conclusion that the applicant was in possession of the claimed species is sufficient.[50] Patents and printed publications in the art should be relied upon to determine whether an art is mature and what the level of knowledge and skill is in the art. In most technologies which are mature, and wherein the knowledge and level of skill in the art is high, a written description question should not be raised for original claims even if the specification discloses only a method of making the invention and the function of the invention.[51] In contrast, for inventions in emerging and unpredictable technologies, or for inventions characterized by factors not reasonably predictable which are known to one of ordinary skill in the art, more evidence is required to show possession. For example, disclosure of only a method of making the invention and the function may not be sufficient to support a product claim other than a product-by-process claim.[52] Furthermore, disclosure of a partial structure without additional characterization of the product may not be sufficient to evidence possession of the claimed invention.[53]

Any claim to a species that does not meet the test described under at least one of (a), (b), or (c) must be rejected as lacking adequate written description under 35 U.S.C. 112, ¶ 1.

(2) For each claim drawn to a genus:

The written description requirement for a claimed genus may be satisfied through sufficient description of a representative number of species by actual reduction to practice (see (1)(a), above), reduction to drawings (see (1)(b), above), or by disclosure of relevant, identifying characteristics, *i.e.*, structure or other physical and/or chemical properties, by functional characteristics coupled with a known or disclosed correlation between function and structure, or by a combination of such identifying characteristics, sufficient to show the applicant was in possession of the claimed genus (see (1)(c), above).[54]

A "representative number of species" means that the species which are adequately described are representative of the entire genus. Thus, when there is substantial variation within the genus, one must describe a sufficient variety of species to reflect the variation within the genus. On the other hand, there may be situations where one species adequately supports a genus.[55] What constitutes a "representative number" is an inverse function of the skill and knowledge in the art. Satisfactory disclosure of a "representative number" depends on whether one of skill in the art would recognize that the applicant was in possession of the necessary common attributes or features of the elements possessed by the members of the genus in view of the species disclosed. For inventions in an unpredictable art, adequate written description of a genus which embraces widely variant species *cannot* be achieved by disclosing only one species within the genus.[56] Description of a representative number of species does not require the description to be of such specificity that it would provide individual support for each species that the genus embraces.[57] If a representative number of adequately described species are not disclosed for a genus, the claim to that genus must be rejected as lacking adequate written description under 35 U.S.C. 112, ¶ 1.

b. New claims, amended claims, or claims asserting entitlement to the benefit of an earlier priority date or filing date under 35 U.S.C. 119, 120, or

365(c). The examiner has the initial burden of presenting evidence or reasoning to explain why persons skilled in the art would not recognize in the original disclosure a description of the invention defined by the claims.[58] However, when filing an amendment an applicant should show support in the original disclosure for new or amended claims.[59] To comply with the written description requirement of 35 U.S.C. 112, ¶ 1, or to be entitled to an earlier priority date or filing date under 35 U.S.C. 119, 120, or 365(c), each claim limitation must be expressly,[60] implicitly,[61] or inherently[62] supported in the originally filed disclosure.[63] Furthermore, each claim must include all elements which applicant has described as essential.[64]

If the originally filed disclosure does not provide support for each claim limitation, or if an element which applicant describes as essential or critical is not claimed, a new or amended claim must be rejected under 35 U.S.C. 112, ¶ 1, as lacking adequate written description, or in the case of a claim for priority under 35 U.S.C. 119, 120, or 365(c), the claim for priority must be denied.

III. Complete Patentability Determination Under All Statutory Requirements and Clearly Communicate Findings, Conclusions, and Their Bases

The above only describes how to determine whether the written description requirement of 35 U.S.C. 112, ¶ 1, is satisfied. Regardless of the outcome of that determination, Office personnel must complete the patentability determination under all the relevant statutory provisions of title 35 of the U.S. Code.

Once Office personnel have concluded analysis of the claimed invention under all the statutory provisions, including 35 U.S.C. 101, 112, 102, and 103, they should review all the proposed rejections and their bases to confirm their correctness. Only then should any rejection be imposed in an Office action. The Office action should clearly communicate the findings, conclusions, and reasons which support them. When possible, the Office action should offer helpful suggestions on how to overcome rejections.

A. For Each Claim Lacking Written Description Support, Reject the Claim Under Section 112, ¶ 1, for Lack of Adequate Written Description

A description as filed is presumed to be adequate, unless or until sufficient evidence or reasoning to the contrary has been presented by the examiner to rebut the presumption.[65] The examiner, therefore, must have a reasonable basis to challenge the adequacy of the written description. The examiner has the initial burden of presenting by a preponderance of evidence why a person skilled in the art would not recognize in an applicant's disclosure a description of the invention defined by the claims.[66] In rejecting a claim, the examiner must set forth express findings of fact regarding the above analysis which support the lack of written description conclusion. These findings should:

(1) Identify the claim limitation at issue; and

(2) Establish a *prima facie* case by providing reasons why a person skilled in the art at the time the application was filed would not have recognized that the inventor was in possession of the invention as claimed in view of the disclosure of the application as filed. A general allegation of "unpredictability in the art" is not a sufficient reason to support a rejection for lack of adequate written description.

When appropriate, suggest amendments to the claims which can be supported by the application's written description, being mindful of the prohibition against the addition of new matter in the claims or description.[67]

B. Upon Reply by Applicant, Again Determine the Patentability of the Claimed Invention, Including Whether the Written Description Requirement Is Satisfied by Reperforming the Analysis Described Above in View of the Whole Record

Upon reply by applicant, before repeating any rejection under 35 U.S.C. 112, ¶ 1, for lack of written description, review the basis for the rejection in view of the record as a whole, including amendments, arguments, and any evidence submitted by applicant. If the whole record now demonstrates that the written description requirement is satisfied, do *not* repeat the rejection in the next Office action. If the record still does not demonstrate that the written description is adequate to support the claim(s), repeat the rejection under 35 U.S.C. 112, ¶ 1, fully respond to applicant's rebuttal arguments, and properly treat any further showings submitted by applicant in the reply. When a rejection is maintained, any affidavits relevant to the 112, ¶ 1, written description requirement,[68] must be thoroughly analyzed and discussed in the next Office action.

Dated: December 29, 2000.

Q. Todd Dickinson,

Under Secretary of Commerce for Intellectual Property and Director of the United States Patent and Trademark Office.

Endnotes

[1] *See, e.g., Vas-Cath, Inc.* v. *Mahurkar*, 935 F.2d 1555, 1560, 19 USPQ2d 1111, 1114 (Fed. Cir. 1991).

[2] *In re Barker*, 559 F.2d 588, 592 n.4, 194 USPQ 470, 473 n.4 (CCPA 1977).

[3] *See Regents of the University of California* v. *Eli Lilly*, 119 F.3d 1559, 1566, 43 USPQ2d 1398, 1404 (Fed. Cir. 1997), *cert. denied*, 523 U.S. 1089 (1998).

[4] *See, e.g., Vas-Cath, Inc.* v. *Mahurkar*, 935 F.2d at 1563, 19 USPQ2d at 1116. Much of the written description case law addresses whether the specification as originally filed supports claims not originally in the application. The issue raised in the cases is most often phrased as whether the original application provides "adequate support" for the claims at issue or whether the material added to the specification incorporates "new matter" in violation of 35 U.S.C. 132. The "written description" question similarly arises in the interference context, where the issue is whether the specification of one party to the interference can support the newly added claims corresponding to the count at issue, *i.e.*, whether that party can "make the claim" corresponding to the interference count. *See, e.g., Martin* v. *Mayer*, 823 F.2d 500, 503, 3 USPQ2d 1333, 1335 (Fed. Cir. 1987).

In addition, early opinions suggest the Patent and Trademark Office was unwilling to find written descriptive support when the only description was found in the claims; however, this viewpoint was rejected. *See In re Koller*, 613 F.2d 819, 204 USPQ 702 (CCPA 1980) (original claims constitute their own description); *accord In re Gardner*, 475 F.2d 1389, 177 USPQ 396 (CCPA 1973); *accord In re Wertheim*, 541 F.2d 257, 191 USPQ 90 (CCPA 1976) (accord). It is now well accepted that a satisfactory description may be in the claims or any other portion of the originally filed specification. These early opinions did not address the quality or specificity of particularity that was required in the description, *i.e.*, how much description is enough.

[5] *Lockwood* v. *American Airlines, Inc.*, 107 F.3d 1565, 1572, 41 USPQ2d 1961, 1966 (Fed. Cir. 1997).

[6] An application specification may show actual reduction to practice by describing testing of the claimed invention or, in the case of biological materials, by specifically describing a deposit made in accordance with 37 CFR 1.801 *et seq. See also Deposit of Biological Materials for Patent Purposes, Final Rule*, 54 FR 34,864 (August 22, 1989) ("The requirement for a specific identification is consistent with the description requirement of the first paragraph of 35 U.S.C. 112, and to provide an antecedent basis for the biological material which either has been or will be deposited before the patent is granted." *Id.* at 34,876. "The description must be sufficient to permit verification that the deposited biological material is in fact that disclosed. Once the

patent issues, the description must be sufficient to aid in the resolution of questions of infringement." *Id.* at 34,880.). Such a deposit is not a substitute for a written description of the claimed invention. The written description of the deposited material needs to be as complete as possible because the examination for patentability proceeds solely on the basis of the written description. *See, e.g., In re Lundak*, 773 F.2d 1216, 227 USPQ 90 (Fed. Cir. 1985). *See also* 54 FR at 34,880 ("As a general rule, the more information that is provided about a particular deposited biological material, the better the examiner will be able to compare the identity and characteristics of the deposited biological material with the prior art.").

[7] *Pfaff* v. *Wells Electronics, Inc.*, 525 U.S. 55, 68, 119 S.Ct. 304, 312, 48 USPQ2d 1641, 1647 (1998); *Eli Lilly*, 119 F.3d at 1568, 43 USPQ2d at 1406.

[8] *See Amgen, Inc.* v. *Chugai Pharmaceutical*, 927 F.2d 1200, 1206, 18 USPQ2d 1016, 1021 (Fed. Cir. 1991) (one must define a compound by "whatever characteristics sufficiently distinguish it").

[9] A description requirement issue can arise for original claims (*see, e.g., Eli Lilly*, 119 F.3d 1559, 43 USPQ2d 1398) as well as new or amended claims. Most typically, the issue will arise in the context of determining whether new or amended claims are supported by the description of the invention in the application as filed (*see, e.g., In re Wright*, 866 F.2d 422, 9 USPQ2d 1649 (Fed. Cir. 1989)), whether a claimed invention is entitled to the benefit of an earlier priority date or effective filing date under 35 U.S.C. 119, 120, or 365(c) (*see, e.g., Tronzo* v. *Biomet, Inc.*, 156 F.3d 1154, 47 USPQ2d 1829 (Fed. Cir. 1998); *Fiers* v. *Revel*, 984 F.2d 1164, 25 USPQ2d 1601 (Fed. Cir. 1993); *In re Ziegler*, 992 F.2d 1197, 1200, 26 USPQ2d 1600, 1603 (Fed. Cir. 1993)), or whether a specification provides support for a claim corresponding to a count in an interference (*see, e.g., Fields* v. *Conover*, 443 F.2d 1386, 170 USPQ 276 (CCPA 1971)).

[10] *Vas-Cath, Inc.* v. *Mahurkar*, 935 F.2d at 1563, 19 USPQ2d at 1116 (Fed. Cir. 1991).

[11] *In re Wertheim*, 541 F.2d 257, 263, 191 USPQ 90, 97 (CCPA 1976) ("we are of the opinion that the PTO has the initial burden of presenting evidence or reasons why persons skilled in the art would not recognize in the disclosure a description of the invention defined by the claims").

[12] See endnote 4.

[13] For example, consider the claim "A gene comprising SEQ ID NO:1." A determination of what the claim as a whole covers may result in a conclusion that specific structures such as a promoter, a coding region, or other elements are included. Although all genes encompassed by this claim share the characteristic of comprising SEQ ID NO:1, there may be insufficient description of those specific structures (*e.g.*, promoters, enhancers, coding regions, and other regulatory elements) which are also included.

[14] A biomolecule sequence described only by a functional characteristic, without any known or disclosed correlation between that function and the structure of the sequence, normally is not a sufficient identifying characteristic for written description purposes, even when accompanied by a method of obtaining the claimed sequence. For example, even though a genetic code table would correlate a known amino acid sequence with a genus of coding nucleic acids, the same table cannot predict the native, naturally occurring nucleic acid sequence of a naturally occurring mRNA or its corresponding cDNA. *Cf. In re Bell*, 991 F.2d 781, 26 USPQ2d 1529 (Fed. Cir. 1993), and *In re Deuel*, 51 F.3d 1552, 34 USPQ2d 1210 (Fed. Cir. 1995) (holding that a process could not render the product of that process obvious under 35 U.S.C. 103). The Federal Circuit has pointed out that under United States law, a description that does not render a claimed invention obvious cannot sufficiently describe the invention for the purposes of the written description requirement of 35 U.S.C. 112. *Eli Lilly*, 119 F.3d at 1567, 43 USPQ2d at 1405.

Compare Fonar Corp. v. *General Electric Co.*, 107 F.3d 1543, 1549, 41 USPQ2d 1801, 1805 (Fed. Cir. 1997) ("As a general rule, where software constitutes part of a best mode of carrying out an invention, description of such a best mode is satisfied by a disclosure of the functions of the software. This is because, normally, writing code for such software is within the skill of the art, not requiring undue experimentation, once its functions have been disclosed. * * * Thus, flow charts or source code listings are not a requirement for adequately disclosing the functions of software.").

[15] *See, e.g., Fujikawa* v. *Wattanasin*, 93 F.3d 1559, 1571, 39 USPQ2d 1895, 1905 (Fed. Cir. 1996) (a "laundry list" disclosure of every possible moiety does not constitute a written description of every species in a genus because it would not "reasonably lead" those skilled in the art to any particular species); *In re Ruschig*, 379 F.2d 990, 995, 154 USPQ 118, 123 (CCPA 1967) ("*If* n-propylamine had been used in making the compound instead of n-butylamine, the compound of claim 13 would have resulted. Appellants submit to us, as they did to the board, an imaginary specific example patterned on specific example 6 by which the above butyl compound is made so that we can see what a simple change would have resulted in a specific supporting disclosure being present in the present specification. The trouble is that there is no such disclosure, easy though it is to imagine it.") (emphasis in original); *Purdue Pharma L.P.* v. *Faulding Inc.*, 230 F.3d 1320, 1328, 56 USPQ2d 1481, 1487 (Fed. Cir. 2000) ("the specification does not clearly disclose to the skilled artisan that the inventors * * * considered the [] ratio to be part of their invention * * *. There is therefore no force to Purdue's argument that the written description requirement was satisfied because the disclosure revealed a broad invention from which the [later-filed] claims carved out a patentable portion").

[16] 35 U.S.C. §§ 132 and 251. *See also In re Rasmussen*, 650 F.2d 1212, 1214, 211 USPQ 323, 326 (CCPA 1981). *See* Manual of Patent Examining Procedure (MPEP) §§ 2163.06–2163.07 (7th Ed., Rev. 1, Feb. 2000) for a more detailed discussion of the written description requirement and its relationship to new matter.

[17] The claims as filed in the original specification are part of the disclosure and, therefore, if an application as originally filed contains a claim disclosing material not found in the remainder of the specification, the applicant may amend the specification to include the claimed subject matter. *In re Benno*, 768 F.2d 1340, 226 USPQ 683 (Fed. Cir. 1985).

[18] *See, e.g., In re Lukach*, 442 F.2d 967, 169 USPQ 795 (CCPA 1971) (subgenus range was not supported by generic disclosure and specific example within the subgenus range); *In re Smith*, 458 F.2d 1389, 1395, 173 USPQ 679, 683 (CCPA 1972) (a subgenus is not necessarily described by a genus encompassing it and a species upon which it reads).

[19] *In re Oda*, 443 F.2d 1200, 170 USPQ 260 (CCPA 1971). With respect to the correction of sequencing errors in applications disclosing nucleic acid and/or amino acid sequences, it is well known that sequencing errors are a common problem in molecular biology. *See, e.g.*, Peter Richterich, *Estimation of Errors in 'Raw' DNA Sequences: A Validation Study*, 8 Genome Research 251–59 (1998). If an application as filed includes sequence information and references a deposit of the sequenced material made in accordance with the requirements of 37 CFR § 1.801 *et seq.*, amendment may be permissible.

[20] Corrections of minor errors in the sequence may be possible based on the argument that one of skill in the art would have resequenced the deposited material and would have immediately recognized the minor error. Deposits made after the filing date can only be relied upon to provide support for the correction of sequence information if applicant submits a statement in compliance with 37 CFR § 1.804 stating that the biological material which is deposited is a biological material specifically defined in the application as filed.

[21] *See, e.g., Gentry Gallery, Inc.* v. *Berkline Corp.*, 134 F.3d 1473, 45 USPQ2d 1498 (Fed. Cir. 1998) (claims to a sectional sofa comprising, *inter alia*, a console and a control means were held invalid for failing to satisfy the written description requirement where the claims were broadened by removing the location of the control means.); *Johnson Worldwide Associates* v. *Zebco Corp.*, 175 F.3d 985, 993, 50 USPQ2d 1607, 1613 (Fed. Cir. 1999) (In *Gentry Gallery*, the "court's determination that the patent disclosure did not support a broad meaning for the disputed claim terms was premised on clear statements in the written description that described the location of a claim element—the 'control means'—as 'the only possible location' and that variations were 'outside the stated purpose of the invention.' *Gentry Gallery*, 134 F.3d at 1479, 45 USPQ2d at 1503. *Gentry Gallery*, then, considers the situation where the patent's disclosure makes crystal clear that a particular (*i.e.*, narrow) understanding of a claim term is an 'essential element of [the inventor's] invention.'"); *Tronzo* v. *Biomet*, 156 F.3d at 1158–59, 47 USPQ2d at 1833 (Fed. Cir. 1998) (claims to generic cup shape were not entitled to filing date of parent application which disclosed "conical cup" in view of the disclosure of the

parent application stating the advantages and importance of the conical shape.).

[22] *See Gentry Gallery*, 134 F.3d at 1480, 45 USPQ2d at 1503; *In re Sus*, 306 F.2d 494, 504, 134 USPQ 301, 309 (CCPA 1962) ("[O]ne skilled in this art would not be taught by the written description of the invention in the specification that any 'aryl or substituted aryl radical' would be suitable for the purposes of the invention but rather that only *certain aryl radicals* and certain specifically substituted aryl radicals (*i.e.*, aryl azides) would be suitable for such purposes.") (emphasis in original). A claim which omits matter disclosed to be essential to the invention as described in the specification or in other statements of record may also be subject to rejection under 35 U.S.C. 112, ¶ 1, as not enabling, or under 35 U.S.C. 112, ¶ 2. *See In re Mayhew*, 527 F.2d 1229, 188 USPQ 356 (CCPA 1976); *In re Venezia*, 530 F.2d 956, 189 USPQ 149 (CCPA 1976); and *In re Collier*, 397 F.2d 1003, 158 USPQ 266 (CCPA 1968). *See also* MPEP § 2172.01.

[23] *See, e.g., Vas-Cath, Inc.*, 935 F.2d at 1563–64, 19 USPQ2d at 1117.

[24] *Wertheim*, 541 F.2d at 262, 191 USPQ at 96.

[25] *See* MPEP §§ 714.02 and 2163.06 ("Applicant should * * * specifically point out the support for any amendments made to the disclosure."); and MPEP § 2163.04 ("If applicant amends the claims and points out where and/or how the originally filed disclosure supports the amendment(s), and the examiner finds that the disclosure does not reasonably convey that the inventor had possession of the subject matter of the amendment at the time of the filing of the application, the examiner has the initial burden of presenting evidence or reasoning to explain why persons skilled in the art would not recognize in the disclosure a description of the invention defined by the claims.").

[26] *See In re Smith*, 458 F.2d 1389, 1395, 173 USPQ 679, 683 (CCPA 1972) ("Precisely how close [to the claimed invention] the description must come to comply with § 112 must be left to case-by-case development."); *In re Wertheim*, 541 F.2d at 262, 191 USPQ at 96 (inquiry is primarily factual and depends on the nature of the invention and the amount of knowledge imparted to those skilled in the art by the disclosure).

[27] *See, e.g., In re Morris*, 127 F.3d 1048, 1053–54, 44 USPQ2d 1023, 1027 (Fed. Cir. 1997).

[28] "Preamble language" is that language in a claim appearing before the transitional phase, *e.g.*, before "comprising," "consisting essentially of," or "consisting of."

[29] The transitional term "comprising" (and other comparable terms, *e.g.*, "containing," "including," and "having") is "open-ended—it covers the expressly recited subject matter, alone or in combination with unrecited subject matter. *See, e.g., Genentech, Inc.* v. *Chiron Corp.*, 112 F.3d 495, 501, 42 USPQ2d 1608, 1613 (Fed. Cir. 1997) ("'Comprising' is a term of art used in claim language which means that the named elements are essential, but other elements may be added and still form a construct within the scope of the claim."); *Ex parte Davis*, 80 USPQ 448, 450 (Bd. App. 1948) ("comprising" leaves the "claim open for the inclusion of unspecified ingredients even in major amounts"). "By using the term 'consisting essentially of,' the drafter signals that the invention necessarily includes the listed ingredients and is open to unlisted ingredients that do not materially affect the basic and novel properties of the invention. A 'consisting essentially of' claim occupies a middle ground between closed claims that are written in a 'consisting of' format and fully open claims that are drafted in a 'comprising' format." *PPG Industries* v. *Guardian Industries*, 156 F.3d 1351, 1354, 48 USPQ2d 1351, 1353–54 (Fed. Cir. 1998). For the purposes of searching for and applying prior art under 35 U.S.C. 102 and 103, absent a clear indication in the specification or claims of what the basic and novel characteristics actually are, 'consisting essentially of' will be construed as equivalent to "comprising." *See, e.g., PPG*, 156 F.3d at 1355, 48 USPQ2d at 1355 ("PPG could have defined the scope of the phrase "consisting essentially of" for purposes of its patent by making clear in its specification what it regarded as constituting a material change in the basic and novel characteristics of the invention."). *See also In re Janakirama-Rao*, 317 F.2d 951, 954, 137 USPQ 893, 895–96 (CCPA 1963). If an applicant contends that additional steps or materials in the prior art are excluded by the recitation of "consisting essentially of," applicant has the burden of showing that the introduction of additional steps or components would materially change the characteristics of applicant's invention. *In re De Lajarte*, 337 F.2d 870, 143 USPQ 256 (CCPA 1964).

[30] *See Pac-Tec Inc.* v. *Amerace Corp.*, 903 F.2d 796, 801, 14 USPQ2d 1871, 1876 (Fed. Cir. 1990) (determining that preamble language that constitutes a structural limitation is actually part of the claimed invention).

[31] An applicant shows possession of the claimed invention by describing the claimed invention with all of its limitations. *Lockwood*, 107 F.3d at 1572, 41 USPQ2d at 1966.

[32] *See, e.g., Bell Communications Research, Inc.* v. *Vitalink Communications Corp.*, 55 F.3d 615, 620, 34 USPQ2d 1816, 1820 (Fed. Cir. 1995) ("[A] claim preamble has the import that the claim as a whole suggests for it."); *Corning Glass Works* v. *Sumitomo Elec. U.S.A., Inc.*, 868 F.2d 1251, 1257, 9 USPQ2d 1962, 1966 (Fed. Cir. 1989) (The determination of whether preamble recitations are structural limitations can be resolved only on review of the entirety of the application "to gain an understanding of what the inventors actually invented and intended to encompass by the claim.").

[33] An element may be critical where those of skill in the art would require it to determine that applicant was in possession of the invention. *Compare Rasmussen*, 650 F.2d at 1215, 211 USPQ at 327 ("one skilled in the art who read Rasmussen's specification would understand that it is unimportant *how* the layers are adhered, so long as they are adhered") (emphasis in original), *with Amgen, Inc.* v. *Chugai Pharmaceutical Co., Ltd.*, 927 F.2d 1200, 1206, 18 USPQ2d 1016, 1021 (Fed. Cir. 1991) ("it is well established in our law that conception of a chemical compound requires that the inventor be able to define it so as to distinguish it from other materials, and to describe how to obtain it").

[34] *See, e.g., Wang Labs.* v. *Toshiba Corp.*, 993 F.2d 858, 865, 26 USPQ2d 1767, 1774 (Fed. Cir. 1993).

[35] *See, e.g., Hybritech, Inc.* v. *Monoclonal Antibodies, Inc.*, 802 F.2d 1367, 1379–80, 231 USPQ 81, 90 (Fed. Cir. 1986).

[36] *See, e.g., Purdue Pharma L.P.* v. *Faulding Inc.*, 230 F.3d 1320, ___, 56 USPQ2d 1481, 1483 (Fed. Cir. 2000) (the written description "inquiry is a factual one and must be assessed on a case-by-case basis"); *see also Pfaff* v. *Wells Electronics, Inc.*, 55 U.S. at 66, 119 S.Ct. at 311, 48 USPQ2d at 1646 ("The word 'invention' must refer to a concept that is complete, rather than merely one that is 'substantially complete.' It is true that reduction to practice ordinarily provides the best evidence that an invention is complete. But just because reduction to practice is sufficient evidence of completion, it does not follow that proof of reduction to practice is necessary in every case. Indeed, both the facts of the *Telephone Cases* and the facts of this case demonstrate that one can prove that an invention is complete and ready for patenting before it has actually been reduced to practice.").

[37] *Cooper* v. *Goldfarb*, 154 F.3d 1321, 1327, 47 USPQ2d 1896, 1901 (Fed. Cir. 1998). *See also UMC Elecs. Co.* v. *United States*, 816 F.2d 647, 652, 2 USPQ2d 1465, 1468 (Fed. Cir. 1987) ("[T]here cannot be a reduction to practice of the invention * * * without a physical embodiment which includes all limitations of the claim."); *Estee Lauder Inc.* v. *L'Oreal, S.A.*, 129 F.3d 588, 593, 44 USPQ2d 1610, 1614 (Fed. Cir. 1997) ("[A] reduction to practice does not occur until the inventor has determined that the invention will work for its intended purpose."); *Mahurkar* v. *C.R. Bard, Inc.*, 79 F.3d 1572, 1578, 38 USPQ2d 1288, 1291 (Fed. Cir. 1996) (determining that the invention will work for its intended purpose may require testing depending on the character of the invention and the problem it solves).

[38] 37 CFR 1.804, 1.809. *See also* endnote 6.

[39] *See, e.g., Vas-Cath*, 935 F.2d at 1565, 19 USPQ2d at 1118 ("drawings alone may provide a 'written description' of an invention as required by § 112"); *In re Wolfensperger*, 302 F.2d 950, 133 USPQ 537 (CCPA 1962) (the drawings of applicant's specification provided sufficient written descriptive support for the claim limitation at issue); *Autogiro Co. of America* v. *United States*, 384 F.2d 391, 398, 155 USPQ 697, 703 (Ct. Cl. 1967) ("In those instances where a visual representation can flesh out words, drawings may be used in the same manner and with the same limitations as the specification.").

[40] *See, e.g., Eli Lilly*, 119 F.3d at 1568, 43 USPQ2d at 1406 ("In claims involving chemical materials, generic formulae usually indicate with specificity what the generic claims encompass. One skilled in the art can distinguish such a formula from others and can identify many of the species that the claims encompass. Accordingly, such a formula is normally an adequate description of the claimed genus.").

[41] *See Hybritech v. Monoclonal Antibodies*, 802 F.2d at 1384, 231 USPQ at 94; *Fonar Corp. v. General Electric Co.*, 107 F.3d at 1549, 41 USPQ2d at 1805 (source code description not required).

[42] For example, the presence of a restriction enzyme map of a gene may be relevant to a statement that the gene has been isolated. One skilled in the art may be able to determine when the gene disclosed is the same as or different from a gene isolated by another by comparing the restriction enzyme map. In contrast, evidence that the gene could be digested with a nuclease would not normally represent a relevant characteristic since any gene would be digested with a nuclease. Similarly, isolation of an mRNA and its expression to produce the protein of interest is strong evidence of possession of an mRNA for the protein.

For some biomolecules, examples of identifying characteristics include a sequence, structure, binding affinity, binding specificity, molecular weight, and length. Although structural formulas provide a convenient method of demonstrating possession of specific molecules, other identifying characteristics or combinations of characteristics may demonstrate the requisite possession. For example, unique cleavage by particular enzymes, isoelectric points of fragments, detailed restriction enzyme maps, a comparison of enzymatic activities, or antibody cross-reactivity may be sufficient to show possession of the claimed invention to one of skill in the art. *See Lockwood*, 107 F.3d at 1572, 41 USPQ2d at 1966 ("written description" requirement may be satisfied by using "such descriptive means as words, structures, figures, diagrams, formulas, etc., that fully set forth the claimed invention").

[43] A definition by function alone "does not suffice" to sufficiently describe a coding sequence "because it is only an indication of what the gene does, rather than what it is." *Eli Lilly*, 119 F.3 at 1568, 43 USPQ2d at 1406. *See also Fiers*, 984 F.2d at 1169–71, 25 USPQ2d at 1605–06 (discussing *Amgen Inc. v. Chugai Pharmaceutical Co.*, 927 F.2d 1200, 18 USPQ2d 1016 (Fed. Cir. 1991)).

[44] If a claim limitation invokes 35 U.S.C. 112, ¶ 6, it must be interpreted to cover the corresponding structure, materials, or acts in the specification and "equivalents thereof." *See* 35 U.S.C. 112, ¶ 6. *See also B. Braun Medical, Inc. v. Abbott Lab.*, 124 F.3d 1419, 1424, 43 USPQ2d 1896, 1899 (Fed. Cir. 1997). In considering whether there is 35 U.S.C. 112, ¶ 1, support for a means- (or step) plus-function claim limitation, the examiner must consider not only the original disclosure contained in the summary and detailed description of the invention portions of the specification, but also the original claims, abstract, and drawings. A means- (or step-) plus-function claim limitation is adequately described under 35 U.S.C. 112, ¶ 1, if: (1) The written description adequately links or associates adequately described particular structure, material, or acts to the function recited in a means- (or step-) plus-function claim limitation; or (2) it is clear based on the facts of the application that one skilled in the art would have known what structure, material, or acts perform the function recited in a means- (or step-) plus-function limitation. Note also: A rejection under 35 U.S.C. 112, ¶ 2, "cannot stand where there is adequate description in the specification to satisfy 35 U.S.C. 112, first paragraph, regarding means-plus-function recitations that are not, per se, challenged for being unclear." *In re Noll*, 545 F.2d 141, 149, 191 USPQ 721, 727 (CCPA 1976). *See Supplemental Examination Guidelines for Determining the Applicability of 35 U.S.C. 112, ¶ 6*, 65 FR 38510, June 21, 2000.

[45] *See Hybritech Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d at 1384, 231 USPQ at 94.

[46] *See, e.g., Vas-Cath*, 935 F.2d at 1563, 19 USPQ2d at 1116; *Martin v. Johnson*, 454 F.2d 746, 751, 172 USPQ 391, 395 (CCPA 1972) (stating "the description need not be in *ipsis verbis* [*i.e.*, "in the same words"] to be sufficient").

[47] A claim which is limited to a single disclosed embodiment or species is analyzed as a claim drawn to a single embodiment or species, whereas a claim which encompasses two or more embodiments or species within the scope of the claim is analyzed as a claim drawn to a genus. *See also* MPEP § 806.04(e).

[48] 35 U.S.C. 112, ¶ 1. *Cf. Fields v. Conover*, 443 F.2d 1386, 1392, 170 USPQ 276, 280 (CCPA 1971) (finding a lack of written description because the specification lacked the "full, clear, concise, and exact written description" which is necessary to support the claimed invention).

[49] For example, if the art has established a strong correlation between structure and function, one skilled in the art would be able to predict with a reasonable degree of confidence the structure of the claimed invention from a recitation of its function. Thus, the written description requirement may be satisfied through disclosure of function and minimal structure when there is a well-established correlation between structure and function. In contrast, without such a correlation, the capability to recognize or understand the structure from the mere recitation of function and minimal structure is highly unlikely. In this latter case, disclosure of function alone is little more than a wish for possession; it does not satisfy the written description requirement. *See Eli Lilly*, 119 F.3d at 1568, 43 USPQ2d at 1406 (written description requirement not satisfied by merely providing "a result that one might achieve if one made that invention"); *In re Wilder*, 736 F.2d 1516, 1521, 222 USPQ 369, 372–73 (Fed. Cir. 1984) (affirming a rejection for lack of written description because the specification does "little more than outline goals appellants hope the claimed invention achieves and the problems the invention will hopefully ameliorate"). *Compare Fonar*, 107 F.3d at 1549, 41 USPQ2d at 1805 (disclosure of software function adequate in that art).

[50] *See Eli Lilly*, 119 F.3d at 1568, 43 USPQ2d at 1406.

[51] *See, e.g., In re Hayes Microcomputer Products, Inc. Patent Litigation*, 982 F.2d 1527, 1534–35, 25 USPQ2d 1241, 1246 (Fed. Cir. 1992) ("One skilled in the art would know how to program a microprocessor to perform the necessary steps described in the specification. Thus, an inventor is not required to describe every detail of his invention. An applicant's disclosure obligation varies according to the art to which the invention pertains. Disclosing a microprocessor capable of performing certain functions is sufficient to satisfy the requirement of section 112, first paragraph, when one skilled in the relevant art would understand what is intended and know how to carry it out.")

[52] *See, e.g., Fiers v. Revel*, 984 F.2d at 1169, 25 USPQ2d at 1605; *Amgen*, 927 F.2d at 1206, 18 USPQ2d at1021. Where the process has actually been used to produce the product, the written description requirement for a product-by-process claim is clearly satisfied; however, the requirement may not be satisfied where it is not clear that the acts set forth in the specification can be performed, or that the product is produced by that process.

[53] *See, e.g., Amgen*, 927 F.2d at1206, 18 USPQ2d at 1021 ("A gene is a chemical compound, albeit a complex one, and it is well established in our law that conception of a chemical compound requires that the inventor be able to define it so as to distinguish it from other materials, and to describe how to obtain it. Conception does not occur unless one has a mental picture of the structure of the chemical, or is able to define it by its method of preparation, its physical or chemical properties, or whatever characteristics sufficiently distinguish it. It is not sufficient to define it solely by its principal biological property, *e.g.*, encoding human erythropoietin, because an alleged conception having no more specificity than that is simply a wish to know the identity of any material with that biological property. We hold that when an inventor is unable to envision the detailed constitution of a gene so as to distinguish it from other materials, as well as a method for obtaining it, conception has not been achieved until reduction to practice has occurred, *i.e.*, until after the gene has been isolated.") (citations omitted). In such instances the alleged conception fails not merely because the field is unpredictable or because of the general uncertainty surrounding experimental sciences, but because the conception is incomplete due to factual uncertainty that undermines the specificity of the inventor's idea of the invention. *Burroughs Wellcome Co. v. Barr Laboratories Inc.*, 40 F.3d 1223, 1229, 32 USPQ2d 1915, 1920 (Fed. Cir. 1994). Reduction to practice in effect provides the only evidence to corroborate conception (and therefore possession) of the invention. *Id.*

[54] *See Eli Lilly*, 119 F.3d at 1568, 43 USPQ2d at 1406.

[55] *See, e.g., Rasmussen*, 650 F.2d at 1214, 211 USPQ at 326–27 (disclosure of a single method of adheringly applying one layer to another was sufficient to support a generic claim to "adheringly applying" because one skilled in the art reading the specification would understand that it is unimportant how the layers are adhered, so long as they are adhered); *In re Herschler*, 591 F.2d 693, 697, 200 USPQ 711, 714 (CCPA 1979) (disclosure of corticosteriod in DMSO sufficient to support claims drawn to a method of using a mixture of a "physiologically active steroid" and DMSO because "use of known chemical compounds in a manner auxiliary

to the invention must have a corresponding written description only so specific as to lead one having ordinary skill in the art to that class of compounds. Occasionally, a functional recitation of those known compounds in the specification may be sufficient as that description."); *In re Smythe*, 480 F.2d 1376, 1383, 178 USPQ 279, 285 (CCPA 1973) (the phrase "air or other gas which is inert to the liquid" was sufficient to support a claim to "inert fluid media" because the description of the properties and functions of the air or other gas segmentizing medium would suggest to a person skilled in the art that appellant's invention includes the use of "inert fluid" broadly.). However, in *Tronzo v. Biomet*, 156 F.3d at 1159, 47 USPQ2d at1833 (Fed. Cir. 1998), the disclosure of a species in the parent application did not suffice to provide written description support for the genus in the child application.

[56] *See, e.g., Eli Lilly.*

[57] For example, in the molecular biology arts, if an applicant disclosed an amino acid sequence, it would be unnecessary to provide an explicit disclosure of nucleic acid sequences that encoded the amino acid sequence. Since the genetic code is widely known, a disclosure of an amino acid sequence would provide sufficient information such that one would accept that an applicant was in possession of the full genus of nucleic acids encoding a given amino acid sequence, but not necessarily any particular species. *Cf. In re Bell*, 991 F.2d 781, 785, 26 USPQ2d 1529, 1532 (Fed. Cir. 1993) and *In re Baird*, 16 F.3d 380, 382, 29 USPQ2d 1550, 1552 (Fed. Cir. 1994).

[58] *See Wertheim*, 541 F.2d at 263, 191 USPQ at 97 ("[T]he PTO has the initial burden of presenting evidence or reasons why persons skilled in the art would not recognize in the disclosure a description of the invention defined by the claims.").

[59] *See* MPEP §§ 714.02 and 2163.06 ("Applicant should * * * specifically point out the support for any amendments made to the disclosure.").

[60] *See, e.g., In re Wright*, 866 F.2d 422, 425, 9 USPQ2d 1649, 1651 (Fed. Cir. 1989) (Original specification for method of forming images using photosensitive microcapsules which describes removal of microcapsules from surface and warns that capsules not be disturbed prior to formation of image, unequivocally teaches absence of permanently fixed microcapsules and supports amended language of claims requiring that microcapsules be "not permanently fixed" to underlying surface, and therefore meets description requirement of 35 U.S.C. 112.).

[61] *See, e.g., In re Robins*, 429 F.2d 452, 456–57, 166 USPQ 552, 555 (CCPA 1970) ("[W]here no explicit description of a generic invention is to be found in the specification * * * mention of representative compounds may provide an implicit description upon which to base generic claim language."); *In re Smith*, 458 F.2d 1389, 1395, 173 USPQ 679, 683 (CCPA 1972) (a subgenus is not necessarily implicitly described by a genus encompassing it and a species upon which it reads).

[62] *See, e.g., In re Robertson*, 169 F.3d 743, 745, 49 USPQ2d 1949, 1950–51 (Fed. Cir. 1999) ("To establish inherency, the extrinsic evidence "must make clear that the missing descriptive matter is necessarily present in the thing described in the reference, and that it would be so recognized by persons of ordinary skill. Inherency, however, may not be established by probabilities or possibilities. The mere fact that a certain thing may result from a given set of circumstances is not sufficient.'") (citations omitted).

[63] When an explicit limitation in a claim "is not present in the written description whose benefit is sought it must be shown that a person of ordinary skill would have understood, at the time the patent application was filed, that the description requires that limitation." *Hyatt v. Boone*, 146 F.3d 1348, 1353, 47 USPQ2d 1128, 1131 (Fed. Cir. 1998).

[64] *See, e.g., Johnson Worldwide Associates Inc. v. Zebco Corp.*, 175 F.3d at 993, 50 USPQ2d at 1613; *Gentry Gallery, Inc. v. Berkline Corp.*, 134 F.3d at 1479, 45 USPQ2d at 1503; *Tronzo v. Biomet*, 156 F.3d at 1159, 47 USPQ2d at 1833.

[65] *See, e.g., In re Marzocchi*, 439 F.2d 220, 224, 169 USPQ 367, 370 (CCPA 1971).

[66] *Wertheim*, 541 F.2d at 263, 191 USPQ at 97.

[67] *See Rasmussen*, 650 F.2d at 1214, 211 USPQ at 326.

[68] *See In re Alton*, 76 F.3d 1168, 1176, 37 USPQ2d 1578, 1584 (Fed. Cir. 1996).

[FR Doc. 01–323 Filed 1–4–01; 8:45 am]
BILLING CODE 3510–16–U

CORPORATION FOR NATIONAL AND COMMUNITY SERVICE

Revision of Currently Approved Information Collection; Comment Request

AGENCY: Corporation for National and Community Service

ACTION: Notice.

SUMMARY: The Corporation for National and Community Service (hereinafter "Corporation"), as part of its continuing effort to reduce paperwork and respondent burden, conducts a preclearance consultation program to provide the general public and Federal agencies with an opportunity to comment on proposed and/or continuing collections of information in accordance with the Paperwork Reduction Act of 1995 (PRA95) (44 U.S.C. 3506(c)(2)(A)). This program helps to ensure that requested data can be provided in the desired format, reporting burden (time and financial resources) is minimized, collection instruments are clearly understood, and the impact of collection requirement on respondents can be properly assessed.

Currently, the Corporation is soliciting comments concerning the proposed revision of its Voucher and Payment Request Form (OMB #3045–0014).

Copies of the forms can be obtained by contacting the office listed below in the address section of this notice.

DATES: Written comments must be submitted to the office listed in the **ADDRESSES** section by March 6, 2001.

ADDRESSES: Send comments to Levon Buller, National Service Trust, Corporation for National and Community Service, 1201 New York Ave., NW., Washington, DC 20525.

FOR FURTHER INFORMATION CONTACT: Levon Buller, (202) 606–5000, ext. 383.

SUPPLEMENTARY INFORMATION: The Corporation is particularly interested in comments which:

- Evaluate whether the proposed collection of information is necessary for the proper performance of the functions of the Corporation, including whether the information will have practical utility;
- Evaluate the accuracy of the agency's estimate of the burden of the proposed collection of information, including the validity of the methodology and assumptions used;
- Enhance the quality, utility and clarity of the information to be collected; and
- Minimize the burden of the collection of information on those who are to respond, including through the use of appropriate automated, electronic, mechanical, or other technological collection techniques or other forms of information technology, *e.g.*, permitting electronic submissions of responses.

Background

The Corporation supports programs that provide opportunities for individuals who want to become involved in national service. The service opportunities cover a wide range of activities over varying periods of time. Upon successfully completing an agreed-upon term of service in an approved AmeriCorps program, a national service participant—an AmeriCorps member—receives an "education award". This award is an amount of money set aside in the member's name in the National Service Trust Fund. This education award can be used to make payments towards qualified student loan or pay for educational expenses at qualified post-secondary institutions and approved school-to-work opportunities programs. Members have seven years in which to draw against any unused balance.

The National Service Trust is the office within the Corporation that administers the education award